Remarks

Reconsideration of this Application is respectfully requested.

Upon entry of the foregoing amendment, claims 26, 28, 33, 117-125, and 127-151 are pending in the application, with claim 26 being the independent claim. Claims 26, 28, and 117-148 have been amended and claims 149-151 have been added. Support for these amendments can be found at least in the cancelled claims and throughout the specification, for example, on page 23, lines 7-20; page 10, table 7; page 99, lines 9-18; and page 101, lines 13-21. It is believed these changes introduce no new matter, and their entry is respectfully requested.

Based on the above amendment and the following remarks, Applicants respectfully request that the Examiner reconsider all outstanding objections and rejections and that they be withdrawn.

Interview

Applicants thank Examiner Nashed for the courteous and helpful interview held with Applicants' undersigned representative on May 23, 2003.

The Sequence Listing

The Office Action, at page 2, maintained that the present application is not in compliance with the sequence rules. The concern expressed in the Office Action is that the specification describes mutations at several specific positions in a protein without describing

the sequence of the protein, and that the sequence in a database may change. (Paper 20, p. 2.) Applicants respectfully disagree with the objection.

Applicants need not disclose art-known or standard amino acid sequences or the positions of conserved residues. As the PTO and the courts have stated, what is known in the art need not be disclosed.

At page 57, the specification describes mutations at positions Asp450 and Asp505 in the RNAse H domain of RSV reverse transcriptase. These positions correspond to conserved residues that were described, for example, in Johnson *et al.*, *PNAS USA 83*:7648-7652 (1986) (IDS Document AT18). Johnson *et al.* published an alignment of RSV, M-MLV, HIV and other reverse transcriptase sequences in 1986, which showed that reverse transcriptases have significant conservation between a 150-residue segment in their carboxyl termini (the RNase H domain) and a 250 residue segment in their amino termini (the polymerase domain). *Id.* p. 7649-50, figures 2-4. Johnson *et al.* also pointed out residues that are conserved between the sequences and identified consensus sequences and motifs. *Id.* One of the sequences aligned in Johnson *et al.* was that of RSV reverse transcriptase. The alignment of the RNAse H domain in Figure 2 of Johnson *et al.* shows both Asp450 and Asp505 of RSV reverse transcriptase, which are marked with asterisks because they are conserved across the aligned sequences. Therefore, the sequence of the RSV reverse transcriptase that was mutanted as described at page 57 need not be disclosed in the sequence listing.

At page 73, the specification makes reference to RSV and AMV reverse transcriptases by their GenBank accession numbers. Assuming, arguendo, that changes in a database sequence are even an issue, the artisan can consult the publications cited in the

databases, see entries for the GenBank accession numbers referred to in the specification[1], or can consult the revision history[2] and the "dead" entries listed therein. Therefore, the sequences of the RSV and AMV reverse transcriptases referred to on page 73 by GenBank accession number need not be disclosed in the sequence listing.

Accordingly, withdrawal of the objection is respectfully requested.

The Rejection Under 35 U.S.C. § 112, First Paragraph Is Traversed

The Office Action, at page 3, maintained the rejection of claims 26, 28, 33, 117-125, and 127-148 under 35 U.S.C. § 112, first paragraph, as allegedly containing subject matter that was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventors had possession of the claimed invention. Applicants respectfully traverse this rejection.

1. The Deposited Biological Material

The Office Action states that the deposited material, if perfected, would overcome the rejection as to claims directed to the wild-type homodimer and heterodimer, and the specific deposited mutants.

Applicants submit herewith a copy of a Declaration Concerning Deposited Biological Material, filed in related Application No. 09/245,026 on October 22, 2002. Applicants therefore respectfully believe all requirements concerning the deposited material have been met. 37 C.F.R. 1.801-1.809. Accordingly, withdrawal of this portion of the rejection is respectfully requested.

---

[1] Copies submitted herewith for the convenience of the Examiner.

[2] Copies submitted herewith for the convenience of the Examiner.

2. The RNase H Domain

The Office Action also states that the specification fails to teach the RNase H domain and residues that affect its activity. Applicants respectfully traverse.

As mentioned above, Johnson *et al.* published an alignment of RSV, M-MLV, HIV and other reverse transcriptase sequences in 1986 and determined that reverse transcriptases have significant conservation between a 150-residue segment in their carboxyl termini, which is the RNase H domain. Johnson, M.S. *et al. PNAS USA 83*:7648-7652 (1986), p. 7649-50, figures 2-4 (IDS Document AT18). Johnson *et al.* also noted the significant conservation between *E. coli* RNAse H and the RNAse H domain of reverse transcriptases. *Id.*, p. 7649, fig. 1. The crystal structures for *E. coli* RNAse H and HIV-1 RNAse H have been determined, confirming that they contain highly related structures. Volkmann, S. *et al.*, *J. Biol. Chem. 268*:2674-2683, p. 2675, col. 1 (1993).[3]

Additionally, Johnson *et al.* pointed out residues within the RNAse H domain that are conserved among the reverse transcriptase sequences they aligned. *Id.* For example, the authors identified residues D450, Q481, E483, L492, N501, D505, S506, H549, N560, and D564 of the RNAse H domain, corresponding to RSV reverse transcriptase, as being identical across the aligned sequences. *Id.*, p. 7649, fig. 2. Others subsequently determined that seven of the amino acids identified by Johnson *et al.* are conserved among retroviral and bacterial sequences and two are invariant among the then known reverse transcriptase sequences. Volkmann *et al.*, p. 2681, col. 1

Moreover, mutations have been made in certain conserved residues that establish which residues are essential for RNAse H activity. The two invariant residues have been

---

[3] Copy enclosed herewith for the convenience of the Examiner.

mutated, creating N494D and Q475E of HIV-1 reverse transcriptase. Volkmann *et al.*, p. 2674, abstract. N494D is similar to the wild-type form in activity. *Id.* This residue is not located in the RNAse H active center, and probably is not involved in substrate binding, but may contribute to tertiary stability. *Id.*, p. 2682, col. 2. On the other hand, Q475E was defective in activity, and probably directly contacts the RNA-DNA substrate. *Id.*, p. 2674, abstract, and p. 2682, col. 2. It has also been determined that D443 and D498 of HIV-1 reverse transcriptase are part of the active center of RNAse H, and that H539 appears to be required for RNA-DNA substrate binding. *Id.*, p. 2675, col. 1.

Therefore, the conserved residues of the RNAse H domain were known, and need not have been disclosed in the present specification. *See Hybritech Inc. v. Monoclonal Antibodies, Inc.,* 802 F.2d at 1384, 231 USPQ at 94; MPEP 2163, p. 2100-165, col. 2 (Rev. 1, Feb. 2003). Accordingly, withdrawal of this portion of the rejection is respectfully requested.

3. Monomeric Form and Multimeric Forms

The Office Action also states that the specification fails to teach a polymerase activity for any form of AMV reverse transcriptase other that a dimer, and that no one has taught a monomeric form of AMV reverse transcriptase. Applicants respectfully traverse.

About 30 years ago, Grandgenett *et al.*, showed that AMV reverse transcriptase is active in the monomeric α form. Grandgenett, D.P., *et al.*, *Proc. Natl. Acad. Sci. USA* *70*:230-234 (1973) (IDS Docuement AT5). The authors stated that the monomeric α form contained both RNA-dependent DNA polymerase activity and RNase H activity. *Id.*, abstract, p. 230. Thus, the Office Action is in error.

With regards to multimeric forms, Applicants have amended the claims to cancel the language "at least one subunit" and "one or more" subunits. Accordingly, withdrawal of this portion of the rejection is respectfully requested.

4. βp4 Subunit

Regarding the term "βp4 subunit," Applicants described the βp4 subunit throughout the specification. As the specification says,

> Various forms of the individual subunits of ASLV RT have been cloned and expressed. These include a 98-kDa precursor polypeptide that is normally processed proteolytically to β and a 4-kDa polypeptide removed from the β carboxy end (Alexander, F., *et al., J. Virol. 61*: 534 (1987) and Anderson, D. *et al., Focus 17*:53 (1995)), and the mature β subunit (Weis, J.H. and Salstrom, J.S., U.S. Patent No. 4, 663, 290 (1987); and Soltis, D.A. and Skalka, A.M., *Proc. Nat. Acad. Sci. USA 85*:3372 (1988)).

Specification, p. 4, lines 22-28. The specification also describes the construction of a gene encoding the mature β subunit by inserting a translational stop codon at the "p4" subunit cleavage site. *See,* p. 56, lines 12-14. As Alexander *et al.* proposed, the β subunit is initially synthesized as a larger precursor that is cleaved to produce the smaller mature form and a 4 kD fragment. Alexander, F., *et al., J. Virol. 61*: 534 (1987), p. 540, figure 6 (IDS Document AT1). Thus, based on the specification and the well known structure of the β subunit precursor, one of ordinary skill in the art would understand that the "βp4 subunit" is the precursor that contains the mature β subunit and the 4 kD fragment. Additionally, mutations in the βp4 subunit are disclosed, for example, at page 20, line 15 to page 22, line 16; and Examples 1 and 7 (pp. 57 and 91-2, and 95). Thus, the specification conveys the structure of the βp4 subunit to one of ordinary skill in the art.

Regarding the activity of the βp4 subunit, the specification shows that dimeric βp4 (βp4βp4) RSV reverse transcriptase does have activity. Specification, p. 102, Table 7. Table 7 shows that the βp4βp4 form has approximately the same level of activity as the ββ form. RSV and AMV are both part of the ASLV, or alphavirus, genus of retroviruses, and their reverse transcriptases have the same subunit and domain structure. *See, e.g.,* Dimcheff, *et al., J. Virology 75*:2002-2009 (2001), p. 2002, col. 1, lines 4-7[4]; The Universal Virus Database of the International Committee on Taxonomy of Viruses, <http://www.ictvdb.iacr.ac.uk/Ictv/fs_retro.htm>[5]; *and* Prasad V.R., "8. Genetic Analysis of Retroviral Reverse Transcriptase Structure and Function," in: *Reverse Transcriptase*, Skalka, A.M. and Goff, S.P., eds., Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 135-162, at p. 135 (1993) (IDS document AR13). One of ordinary skill would recognize evidence of the activity of the βp4βp4 form of RSV reverse transcriptase as being evidence of the activity of the βp4βp4 form of AMV reverse transcriptase. Accordingly, withdrawal of this portion of the rejection is respectfully requested.

The Rejections Under 35 U.S.C. § 112, Second Paragraph Are Traversed

The Office Action, at page 4, maintained the rejection of claims 26, 28, 33, 40, 117-125, and 127-148 under 35 U.S.C. § 112, second paragraph as allegedly being indefinite. Applicants respectfully traverse this rejection.

With regards to the recited "polymerase activity," Applicants have amended the claims as suggested by the Examiner. Accordingly, this portion of the rejection is moot.

---

[4] Copy enclosed herewith for the convenience of the Examiner.

[5] Copy enclosed herewith for the convenience of the Examiner.

With regards to the recitation of "one or more subunits," the Office Action is mistaken in the statement that no monomers of AMV reverse transcriptase have activity. As discussed above, Grandgenett *et al.* found that the α monomeric form of AMV reverse transcriptase has RNA-dependent DNA polymerase activity. Nevertheless, Applicants have amended the claims to cancel the language the Examiner found objectionable, merely to expedite prosecution. Accordingly, this portion of the rejection is moot.

With regards to the term "βp4 subunit," the Office Action is correct in the statement that this term refers to the precursor that produces the β subunit. However, the Office Action is incorrect in the statement that the β subunit is formed by cleavage from the α subunit. In fact, the opposite is true: the α subunit is formed by cleavage from the β precursor subunit. As the specification states, ALSV reverse transcriptase is "a heterodimer of two subunits, α (approximately 62 kDa) and β (approximately 94 kDa), in which α is derived from β by proteolytic cleavage." Specification, p. 3, line 27 to p. 4, line 10.

The specification also states:

> Various forms of the individual subunits of ASLV RT have been cloned and expressed. These include a 98-kDa precursor polypeptide that is normally processed proteolytically to β and a 4-kDa polypeptide removed from the β carboxy end (Alexander, F., *et al., J. Virol. 61*: 534 (1987) and Anderson, D. *et al., Focus 17*:53 (1995)), and the mature β subunit (Weis, J.H. and Salstrom, J.S., U.S. Patent No. 4,663, 290 (1987); and Soltis, D.A. and Skalka, A.M., *Proc. Nat. Acad. Sci. USA 85*:3372 (1988)).

Specification, p. 4, lines 22-28. The specification also describes the construction of a gene encoding the mature β subunit by inserting a translational stop codon at the "p4" subunit cleavage site. *See* specification, p. 56, lines 12-14. As Alexander *et al.* proposed, the β subunit is initially synthesized as a larger precursor that is cleaved to produce the mature

form and a 4 kD fragment. Alexander, F., *et al., J. Virol. 61*: 534 (1987), p. 540, figure 6 (IDS Document AT1). Thus, based on the specification and the well known structure of the β subunit precursor, one of ordinary skill in the art would understand that the "βp4 subunit" is the precursor that contains the mature β subunit and the 4 kD fragment. Thus, the term is clear and definite to one of ordinary skill in the art. Accordingly, withdrawal of this portion of the rejection is respectfully requested.

The Rejection Under 35 U.S.C. § 102(b) Is Traversed

The Office Action, at page 6, maintained the rejection of claims 26, 28, 33, 40, 117-119, 121-125, and 127-148 under 35 U.S.C. § 102(b) as allegedly being anticipated by Soltis *et al.* (*Proc. Natl. Acad. Sci. USA 85*:3372-76 (1988)). Applicants respectfully traverse this rejection.

The Office Action stated that the rejection would be withdrawn if the claims were amended to recite "RNA-dependent DNA" polymerase specific activity. Applicants have amended the claims as suggested by the Examiner. Accordingly, withdrawal of the rejection is respectfully requested.

The Rejection Under 35 U.S.C. § 103(a) Is Traversed

The Office Action, at pages 7-8, maintained the rejection of claims 26, 28, 33, 40, 117-119, 121-125, and 127-148 under 35 U.S.C. § 103(a) as allegedly being unpatentable over Soltis *et al.* in view of the state of the art at the time the application was filed. Applicants respectfully traverse this rejection.

To establish a prima facie case of obviousness, the prior art must teach or suggest all the claim limitations and there must be a reasonable expectation of success. *Smiths Indus. Med. Sys. v. Vital Signs, Inc.*, 183 F.3d 1347, 1356, 51 USPQ2d 1415, 1420 (Fed. Cir. 1999). The prior art *as a whole* must be considered in an obviousness analysis. *In re Young*, 927 F.2d 588, 10 USPQ2d 1089 (Fed. Cir. 1991); MPEP 2143.01, p. 2100-126, col. 1. Moreover, the mere fact that references could be combined, or that the modifications were within the level of ordinary skill, is irrelevant absent an objective motivation to combine or modify the teachings. *In re Mills*, 916 F.3d 680 (Fed. Cir. 1990); *In re* Kotszab, 217 F.3d 1365, 1371 (Fed. Cir. 2000); Ex *parte Levengood*, 28 USPQ2d 1300 (Bd. Pat. App. & Inter. 1993).

Soltis *et al.* disclose the independent expression in *E. coli* of the AMV reverse transcriptase α and β subunits, and the purification of each to a specific activity 10,000-100,000 fold *lower* than the specific activity of AMV reverse transcriptase purified directly from AMV. Soltis et al. also disclose the expression and purification of the full length *pol* product in *E. coli*.

The pending claims are directed to a method of producing AMV reverse transcriptase by expressing it in a eukaryotic cell, and isolating or purifying the expressed reverse transcriptase, with the resulting AMV reverse transcriptase having a specific activity of at least about 30,000 units per milligram. Thus, the present claims recite (1) expression in eukaryotic cells, (2) of AMV enzyme, (3) having a specific activity of ≥30,000 units per milligram. Soltis *et al.* do not teach or suggest all three limitations. Further, there was no motivation to modify Soltis *et al.* based on knowledge in the art as exemplified by Ford *et al.*, *Prot. Express. Purif. 2*:95-107 (1991), as cited by the Examiner.

The priority application for the present application was filed April 22, 1997. In 1997, the artisan had no motivation to produce recombinant AMV reverse transcriptase because *non-recombinant* AMV reverse transcriptase with a very high specific activity was already available by 1995. The Boehringer Mannheim catalog, pages 92-93 (1995) (IDS Document AR17), lists an AMV reverse transcriptase *purified from AMV* that has a specific activity of >50,000 units per milligram.

Moreover, Applicants assert that when the art *as a whole* is considered, not only would there have been no motivation to lead the artisan of ordinary skill to modify Soltis *et al.* to obtain the presently claimed invention, but there would not have been a reasonable expectation of successfully obtaining the presently claimed invention. Unlike the situation for other reverse transcriptases, there were difficulties associated with attempts to produce AMV reverse transcriptase (and other ASLV reverse transcriptases) via recombinant methods. For example, Prasad, V.R., in a review article, stated,

> The avian retroviral RT was the first enzyme to be purified to homogeneity and characterized biochemically; it has become the staple of many recombinant DNA procedures because it is readily available in very high quantities from avian myeloblastosis virus (AMV) particles. However, the genetics of the avian RT has lagged behind that of the mammalian enzymes. It is, in general, more difficult to transfect and clone avian cells than mammalian cells, and thus it has been harder to generate mutant producer cell lines. *The more complex subunit structure of the avian RT has also inhibited its preparation through recombinant DNA methods.* Despite these difficulties, considerable information has been obtained concerning this RT.

Prasad, paragraph spanning pp. 138-139 (emphasis added) (IDS document AR13). Thus, even though reverse transcriptase from AMV was isolated and characterized before any mammalian reverse transcriptases, it was not produced at high specific activity levels by

recombinant methods until the present invention was made – well after mammalian enzymes such as MMLV reverse transcriptase and HIV-1 reverse transcriptase, which were first produced recombinantly in 1985 and 1987, respectively. Prasad, page 140, first paragraph, and page 144, second paragraph.

Further, the art actually taught away from the claimed invention. For example, Kawa, S. *et al., Prot. Expression and Purification 4*:298 (1993) (IDS document AR8) disclose the expression of HIV reverse transcriptase in insect cells to obtain an enzyme with a specific activity of 103 units per milligram.[6] The authors also report that HIV reverse transcriptase produced in *E. coli* has a specific activity of 160 units per milligram. Thus, the method taught in Kawa *et al.*, using expression in insect cells, produced an enzyme with lower specific activity than a method using expression in *E. coli*.

Likewise, the art discloses that HIV-1 reverse transcriptase produced in yeast has a lower specific activity than that produced in *E. coli*. For example, Barr *et al.* produced HIV-1 reverse transcriptase in yeast, and obtained enzyme with a specific activity of 34,666 units per milligram. Barr, P.J. *et al., BioTechnology 5*:486 (1987) (IDS document AT2). Lowe *et al.* produced HIV-1 reverse transcriptase in *E. coli,* and obtained enzyme with a specific activity of 1,610 units per milligram. Lowe, D.M. *et al., Biochemistry 27*:8884 (1988) (IDS document AR11). However, Barr *et al.* and Lowe *et al.* used different definitions for

---

[6] A unit is defined therein as the "amount of enzyme required to catalyze the incorporation of 1 nmol of dTMP in *1 minute* at 25°C using poly(A)-oligo$(dT)_{16}$ as template-primer." p. 302, first paragraph (emphasis added). This is the equivalent of 1030 units per milligram, when a unit is defined as the amount enzyme required to catalyze the incorporation of 1 nmol of dTMP in *10 minutes* at 25°C. At 37°C, the specific activity would be higher.

specific activity.[7] When the specific activity of each is adjusted to define a unit as the amount of enzyme required to incorporate 1nmol dTMP at 37°C in *10 minutes*, as defined in the present specification, then the following results were obtained by Barr *et al.* and Lowe *et al.*:

Host cell	Specific Activity (units per milligram)
Yeast	5,778
E. coli	16,1000

Therefore, the art taught away from expressing reverse transcriptases in eukaryotic cells such as insect cells or yeast cells to obtain a reverse transcriptase with high specific activity.

In summary, Applicants respectfully submit that there was no motivation to produce recombinant AMV reverse transcriptase because non-recombinant enzyme having very high specific activity was already available. Further, even if there had been a motivation to produce recombinant AMV reverse transcriptase, there was no reasonable expectation of successfully obtaining recombinant enzyme with high specific activity because of the complex structure of the AMV enzyme. Moreover, the art taught away from the present claims, which recite a method of production in eukaryotic cells to obtain an AMV enzyme having a specific activity of at least about 30,000 units per milligram. Therefore, the claimed invention is non-obvious over the prior art.

---

[7] Barr *et al.* defined a unit as the amount of enzyme required to incorporate 1 nmol of dTMP in *60 minutes* at 37°C. Barr *et al.*, page 488, table 1. Lowe *et al.* defined a unit as the amount of enzyme required to incorporate 1 nmol of dTMP in *1 minute* at 37°C. Lowe *et al.*, page 8886, table I.

During the interview, the Examiner questioned the applicability of the specific activity of RSV reverse transcriptase described in the specification to the claims, which recite AMV reverse transcriptase. As discussed above on page 16, one of ordinary skill would recognize evidence concerning RSV reverse transcriptase as being evidence concerning AMV reverse transcriptase. Nonetheless, Applicants also provide Gerard, G.F *et al.*, *Nucleic Acids Research 30*:3118-3129 (2002)[8] as additional evidence regarding the specific activity of AMV reverse transcriptase produced by the method of the invention. On page 3119, second column, the authors describe the purification and characterization of AMV reverse transcriptase produced in insect cells. As noted, the recombinant enzyme had a specific activity of 57,500 units per milligram. Thus, the claimed method produces both RSV reverse transcriptase and AMV reverse transcriptase having a specific activity of at least about 30,000 units per milligram.

Applicants therefore respectfully assert that the present method is non-obvious over the prior art. Accordingly, withdrawal of this rejection is respectfully requested.

Conclusion

All of the stated grounds of objection and rejection have been properly traversed, accommodated, or rendered moot. Applicants therefore respectfully request that the Examiner reconsider all presently outstanding objections and rejections and that they be withdrawn. Applicants believe that a full and complete reply has been made to the outstanding Office Action and, as such, the present application is in condition for allowance.

---

[8] Copy submitted herewith for the convenience of the Examiner.

If the Examiner believes, for any reason, that personal communication will expedite prosecution of this application, the Examiner is invited to telephone the undersigned at the number provided.

Prompt and favorable consideration of this Amendment and Reply is respectfully requested.

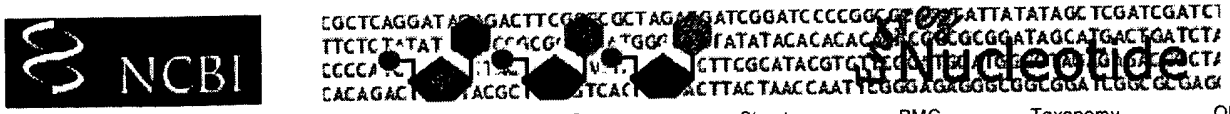
Respectfully submitted,

STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

Helene C. Carslon
Agent for Applicants
Registration No. 47,473

Date: July 21, 2003
1100 New York Avenue, N.W.
Washington, D.C. 20005-3934
(202) 371-2600

SKGF_DC1:129962.5



Links

☐ **1:** L10922. Myeloblastosis-as...[gi:407983]

```
LOCUS       MAVTYP1AAA                7708 bp    DNA     linear   VRL 21-NOV-2002
DEFINITION  Myeloblastosis-associated virus type 1 env gene, complete cds; gag
            gene, complete cds; pol gene, 3' end of cds.
ACCESSION   L10922
VERSION     L10922.1  GI:407983
KEYWORDS    env protein; gag protein; long terminal repeat; pol protein.
SOURCE      Avian myeloblastosis-associated virus type 1
  ORGANISM  Avian myeloblastosis-associated virus type 1
            Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
REFERENCE   1  (bases 1 to 7708)
  AUTHORS   Joliot,V., Boroughs,K., Lasserre,F., Crochet,J., Dambrine,G.,
            Smith,R.E. and Perbal,B.
  TITLE     Pathogenic potential of myeloblastosis-associated virus:
            implication of env proteins for osteopetrosis induction
  JOURNAL   Virology 195 (2), 812-819 (1993)
  MEDLINE   93331743
   PUBMED   8393249
FEATURES             Location/Qualifiers
     source          1..7708
                     /organism="Avian myeloblastosis-associated virus type 1"
                     /proviral
                     /mol_type="genomic DNA"
                     /isolation_source="MAV-1 infected chicken embryo
                     fibroblast"
                     /db_xref="taxon:31669"
     LTR             1..386
     repeat_unit     1..288
                     /note="U3 repeat"
     repeat_unit     289..308
                     /note="R repeat"
     repeat_unit     309..386
                     /note="U5 repeat"
     gene            668..7162
                     /gene="env"
     CDS             join(668..685,5366..7162)
                     /gene="env"
                     /codon_start=1
                     /protein_id="AAA46303.1"
                     /db_xref="GI:407985"
                     /translation="MEAVIKAFLTGRPGKASKKDSKKKPPATSKKDPEKTPLLPTRVN
                     YILIIGVLVLCEVTGVRADVHLLEQPGNLWITWANRTGQTDFCLSTQSATSPFQTCLI
                     GIPSPISEGDFKGYVSDTNCTTLGTDRLVSSASITGGPDNSTTLTYRKVSCLLLKLNV
                     SMWDEPPELQLLGSQSLPNITNITQISGVTGGCVGFRPGGIPWYLGWTRQEATRFLLR
                     QSSFSNSTEPFTVVTADRHNLFMGSEYCGAYGYRFWEVYNCSQEQGQGQQQYRCGKAR
                     SPRPGSPETQCIRRGGKWVNQSQEINETEPFSFTVNCTASNLRNASGCCGKTGMILPG
                     AWIDSTKGNFTKPKALPPGIFLICGDRAWQGIPSRPVGGPCYLGKLTMLAPNHTDILK
                     ILVNSSRTGIRRKRSVSHLDDTCSDEVQLWGPTARIFASILAPGVAAAQALREIERLA
                     CWSVKQANLTTSLLGDLLDDVTSIRHAVLQNRAAIDFLLLAHGHGCEDVAGMCCFNLS
```

```
                     DHSESIQKKFQLMKEHVSKIGVDSDPIGSWLRGLFGGIGEWAVHLLKGLLLGLVVILL
                     LVVCLPCLLQFVSSSIRRIINSSISYHTEYRKMQGGAV"
     gene            668..2773
                     /gene="gag"
     CDS             668..2773
                     /gene="gag"
                     /codon_start=1
                     /protein_id="AAA46302.1"
                     /db_xref="GI:407986"
                     /translation="MEAVIKVISSACKTYCGKTSPSKKEIGAMLSLLQKEGLLMSPSD
                     LYSPGSWDPITAALSQRAMVLGKSGELKTWGLVLGALKAAREEQVTSEQAKFWLGLGG
                     GRVSPPGPECIEKPATERRTDKGEEVGETTAHGDAKMAPEETATPKTVGTSCYHCGTA
                     IGCSCATASAPPPPYVGSGLYPSLAGVGEQQGQEGDTPRGAEQPRAEPGHAGLAPGPA
                     LTDWARVREELASTGPPVVAMPVVIKTEGPAWTPLEPKLITRLADTVRTKGLRSPITM
                     AEVEALMSSPLLPHDVTNLMRVILGPAPYALWMDAWGVQLQTVIAAATRDPRHPANGQ
                     GRGERTNLDRLKGLADGMVGNPQGQAALLRPGELVAITAAALQAFREVARLAEPAGPW
                     ADITQGPSEPFVDFANRLIKAVEGSDLPPSARAPVIIDCFRQKSQPDIQQLIRAAPST
                     LTTPGEIIKYVLDRQKTAPLTDQGIAAAMSSAIQPLVMAVVNRERDGQTGSGGRARGL
                     CYTCGSPGHYQAQCPKKRKSGNSRERCQLCDGMGHNAKQCRKRDGNQGQRPGRGLSSG
                     PWPVSEPPAVSLAMTMEHKDRPLVRVILTNTGSHPVKQRSVYITALLDSGADITIISE
                     EDWPADWPVMEAANPQIHGIGGGIPMRKSRDMIEVGVINRDGSLERPLLLFPAVAMVR
                     GSILGRDCLQGLGLRLTNL"
     gene            2775..5478
                     /gene="pol"
     CDS             <2776..5478
                     /gene="pol"
                     /note="5' end undetermined."
                     /codon_start=1
                     /protein_id="AAA46304.1"
                     /db_xref="GI:407984"
                     /translation="RATVLTVALHLAIPLKWKPNHTPVWIDQWPLPEGKLVALTQLVE
                     KELQLGHIEPSLSCWNTPVFVIRKASGSYRLLHDLRAVNAKLVPFGAVQQGAPVLSAL
                     PRGWPLMVLDLKDCFFSIPLAEQDREAFAFTLPSVNNQAPARRFQWKVLPQGMTCSPT
                     ICQLIVGQILEPLRLKHPSLRMLHYMDDLLLAASSHDGLEAAGEEVISTLERAGFTIS
                     PDKVQREPGVQYLGYKLGSTYVAPVGLVAEPRIATLWDVQKLVGSLQSVRPALGIPPR
                     LMGPFYEQLRGSDPNEAREWNLDMKMAWREIVQLSTTAALERWDPALPLEGAVARCEQ
                     GAIGVLGQGLSTHPRPCLWLFSTQPTKAFTAWLEVLTLLITKLRASAVRTFGKEVDIL
                     LLPACFREDLPLPEGILLALRGFAGKIRSSDTPSIFDIARPLHVSLKVRVTDHPVPGP
                     TVFTDASSSTHKGVVVWREGPRWEIKEIADLGASVQQLEARAVAMALLLWPTTPTNVV
                     TDSAFVAKMLLKMGQEGVPSTAAAFILEDALSQRSAMAAVLHVRSHSEVPGFFTEGND
                     VADSQATFQAYPLREAKDLHTALHIGPRALSKACNISMQQAREVVQTCPHCNSAPALE
                     AGVNPRGLGPLQIWQTDFTLEPRMAPRSWLAVTVDTASSAIVVTQHGRVTSVAAQHHW
                     ATAIAVLGRPKAIKTDNGSCFTSKSTREWLARWGIAHTTGIPGNSQGQAMVERANRLL
                     KDKIRVLAEGDGFMKRIPTSKQGELLAKAMYALNHFERGENTKTPIQKHWRPTVLTEG
                     PPVKIRIETGEWEKGWNVLVWGRGYAAVKNRDTDKVIWVPSRKVKPDITQKDEVTKKD
                     EASPLFAGISDWAPWEGEQEGLQEETASNKQERPGEDTPAANES"
BASE COUNT     1873 a   1868 c   2204 g   1763 t
ORIGIN
        1 tgtagtctta atcgtaggtt aacatgtata ttaccaaata agggaatcgc ctgatgcacc
       61 aaataaggta ttatatgatc ccattggtgg tgaaggagcg acctgagggc atatgggcgt
      121 taacagaact gtctgtcctt gcgtcattcc tcatcggatc atgtacgcgg cagagtatga
      181 ttggataaca ggatggcacc attcatcgtg gcgcatgctg attggtgcga ctaaggagtt
      241 gtgtaaccca cgaatgtact taagcttgta gttgctaaca ataaagtgcc attctacctc
      301 tcaccacatt ggtgtgcacc tgggttgatg gccggaccgt cgattccctg acgactgcga
      361 acacctgaat gaagctgaag gcttcatttg gtgaccccga cgtgatcgtt agggaatagt
      421 ggtcggccac aggcggcgtg gcgatcctgt ctccatccgt ctcgtcgacc gggaggcgag
      481 ttcgatgacc ctgatagagg gggctgcggc ttagggaggc agaagctgag caccgtcggt
      541 aggagctcca gggcccggag cggatacccc tgccgagaac tcagacagtc gtcggaagac
      601 gcggagggag cccgacgatt gagtggtccg ttgttctgcg cgattccggt cgcccggtgg
      661 atcaaacatg gaagccgtca taaaggtgat ttcgtccgcg tgtaaaacct attgcgggaa
      721 aacctctcct tctaagaagg aaataggggc catgttgtcc ctgttacaaa aggaagggtt
```

```
 781 gcttatgtcc ccctcagact tatactcccc ggggtcctgg gatcctatta ccgcggcgct
 841 ctcccagcga gctatggtac ttgggaaatc gggagagtta aaaacctggg gattggtttt
 901 gggggcattg aaggcagctc gagaggaaca ggttacatct gagcaagcaa agttttggtt
 961 gggattaggg ggagggaggg tctctccccc aggtccggag tgtatcgaga aaccagcaac
1021 agagcggcga accgacaagg gggaggaagt gggagaaaca actgcgcacg gagatgcgaa
1081 gatggcgccg gaggaaacgg ccacacctaa gaccgttggc acatcctgct atcattgcgg
1141 aacagctatt ggctgtagtt gcgccacagc ctcggctcct cctcctcctt atgtggggag
1201 tggtttgtat ccttccctgg cgggggtggg agagcagcag ggccaggagg gtgacacacc
1261 tcgggggggcg gaacagccaa gggcggagcc agggcacgcg ggtctggccc ctgggccggc
1321 cctgactgat tgggcaaggg tcagggagga gcttgcgagt actggtccgc ccgtggtggc
1381 catgcctgta gtgattaaga cagagggacc tgcctggacc cctctggagc caaaattgat
1441 cacaagactg gctgatacgg tcaggaccaa gggcttacga tccccgatca ctatggcaga
1501 agtggaagcg cttatgtcct ccccgctgct gccgcatgac gttacgaatc taatgagagt
1561 tattttagga cctgccccat atgccttatg gatggacgct tggggagtcc aattacagac
1621 ggttatagcg gcagccactc gcgacccccg acacccagcg aacggtcaag ggcgggggga
1681 acggactaac ttggatcgct taaagggttt agctgatggg atggtgggca acccacaggg
1741 tcaggccgca ttattaagac cgggggaatt ggttgctatt acggcggcgg ctctccaggc
1801 gtttagagag gttgcccggc tggcggaacc tgcaggtccg tgggcggata tcacgcaggg
1861 accatctgag ccctttgttg actttgccaa tcggcttata aaggcggttg aggggtcaga
1921 tctcccgcct tccgcgcggg ctccggtgat cattgactgc tttaggcaga agtcacagcc
1981 agatatccag cagcttatac gggcagcacc ctccacgctg accaccccag gagagataat
2041 caaatatgtg ctagacaggc agaagactgc ccctcttacg gatcaaggca tagccgcggc
2101 catgtcgtct gctatccagc ccttagttat ggcagtagtc aatagagaga gggatggaca
2161 aactgggtcg ggtggtcgtg cccgagggct ctgctacact tgtggatccc cgggacatta
2221 tcaggcgcag tgcccgaaaa aacggaagtc aggaaacagc cgtgagcgat gtcagttgtg
2281 tgacgggatg ggacacaacg ctaaacagtg taggaagcgg gatggcaacc agggccaacg
2341 cccaggaaga ggcctctctt cggggccgtg gcccgtctct gagccgcctg ccgtctcgtt
2401 agcgatgaca atggagcata aagatcgccc cttggttagg gtcattctga ctaacactgg
2461 gagtcatccg gtcaaacagc gttcggtgta tatcaccgcg ctgttggact ctggagcgga
2521 catcactatt atttcagagg aggactggcc cgccgattgg ccagtgatgg aggctgcgaa
2581 cccacagatc catgggatag gaggggggaat tcccatgcga aaatcccggg atatgataga
2641 agtgggggtt attaaccgag acggatcctt ggagcgaccc ctgctcctct tccccgcagt
2701 ggctatggtt agagggagca tcctaggaag agattgtctg cagggcctag ggctccgctt
2761 gacaaattta tagggagggc cactgttctt actgttgcgc tacatctggc tattccgctc
2821 aaatggaagc caaaccacac gcctgtgtgg attgaccagt ggccccttcc tgaaggtaaa
2881 cttgtagcgc taacgcaatt agtggaaaaa gaattacagt taggacatat agaaccttca
2941 cttagttgct ggaacacacc tgtctttgtg atccggaagg cttccgggtc ttatcgctta
3001 ttgcatgact tgcgcgctgt taacgctaag cttgttcctt ttggggccgt ccaacagggg
3061 gcgccggttc tctccgcgct cccgcgtggt tggcccctga tggtcctaga cctcaaggat
3121 tgcttctttt ctattcctct tgcggaacaa gatcgcgaag cttttgcatt tacgctcccc
3181 tccgtgaata accaggcccc cgctcgaagg ttccaatgga aggtcttgcc ccaagggatg
3241 acctgttctc ccactatctg tcagttgata gtgggtcaaa tacttgagcc cttgcgactc
3301 aagcacccat ctctgcgcat gttgcattat atggatgatc ttttgctagc cgcctcaagt
3361 catgatgggt tggaagcggc aggggaggag gttatcagta cattggaaag agccgggttc
3421 accatttcgc ctgataaggt ccagagggag cccggagtac aatatcttgg gtacaagtta
3481 ggtagtacgt atgtagcacc cgtaggcctg gtagcagaac ccaggatagc caccttgtgg
3541 gatgttcaga agctggtggg gtcacttcag tccgttcgcc cagcgttagg aatcccgcca
3601 cgactgatgg gccccttta tgagcagtta cgagggtcag atcctaacga ggcgagggaa
3661 tggaatctag acatgaaaat ggcctggaga gagatcgtac agctcagcac cactgctgcc
3721 ttggagcgat gggaccctgc cctgcctctg gaaggagcgg tcgctagatg tgaacagggg
3781 gcaatagggg tcctgggaca gggactgtcc acacacccaa ggccatgttt gtggctattc
3841 tccacccaac ccaccaaggc gtttactgct tggttagaag tgctcaccct tttgattact
3901 aagctacgtg cttcggcagt gcgaaccttt ggcaaggagg ttgatatcct cctgttgcct
3961 gcatgctttc gggaggacct tccgctcccg gaggggatcc tgttagccct taggggttt
4021 gcaggaaaaa tcaggagtag tgacacgcca tctatttttg acattgcgcg tccactgcat
4081 gtttctctga aagtgagggt taccgaccac cctgtaccgg gacccactgt ctttactgac
4141 gcctcctcaa gcacccataa gggggtggta gtctggaggg agggcccaag gtgggagata
4201 aaagaaatag ctgatttggg ggcaagtgta caacaactgg aagcacgcgc tgtggccatg
4261 gcacttctgc tgtggccgac aacgcccact aatgtagtga ctgactctgc gtttgttgcg
4321 aaaatgttac tcaagatggg acaggaggga gtcccgtcta cagcggcggc tttcatttta
4381 gaggatgcgt taagccaaag gtcagccatg gccgccgttc tccacgtgcg gagtcattct
```

```
4441 gaagtgccag ggtttttcac agaaggaaat gacgtggcag atagccaagc cacctttcaa
4501 gcgtatccct tgagagaggc taaagatctc cataccgctc tccatattgg accccgcgcg
4561 ctatccaaag cgtgtaatat atctatgcag caggctaggg aggttgttca gacctgcccg
4621 cattgtaatt cagcccctgc gttggaggcc ggggtaaacc ctaggggttt gggacccctа
4681 cagatatggc agacagactt tacacttgag cctagaatgg ctccccgttc ctggctcgct
4741 gttactgtgg ataccgcctc atcggcgata gtcgtaactc agcatggccg tgtcacatcg
4801 gttgctgcac aacatcattg ggccacggct atcgccgttt tgggaagacc aaaggccata
4861 aaaacagata atgggtcctg cttcacgtct aaatccacgc gagagtggct cgcgagatgg
4921 gggatagcac acaccaccgg gattccgggt aattcccagg gtcaagctat ggtagagcgg
4981 gccaaccggc tcctgaaaga taagatccgt gtgcttgcgg agggggatgg ctttatgaaa
5041 agaatcccca ccagcaaaca gggggaacta ttagccaagg caatgtatgc cctcaatcac
5101 tttgagcgtg gtgaaaacac aaaaacaccg atacaaaaac actggagacc taccgttctt
5161 acagaaggac ccccggttaa aatacgaata gagacagggg agtgggaaaa aggatggaac
5221 gtgctggtct ggggacgagg ttatgccgct gtgaaaaaca gggacactga taaggttatt
5281 tgggtaccct ctcgaaaagt taaaccggac atcacccaaa aggatgaggt gactaagaaa
5341 gatgaggcga gccctctttt tgcaggcatt tctgactggg cgccctggga aggcgagcaa
5401 gaaggactcc aagaagaaac cgccagcaac aagcaagaaa gacccggaga agacacccct
5461 gctgccaacg agagttaatt atattctcat tattggtgtc ttggtcttgt gtgaggttac
5521 gggggtaaga gctgatgtcc acttactcga gcaaccaggg aacctttgga ttacatgggc
5581 caaccgtaca ggccaaacgg atttctgcct ctctacacag tcggccacct ccccttttca
5641 aacatgtttg ataggtatcc cgtcccctat ttccgaaggt gattttaagg gatatgtctc
5701 tgatacaaat tgcaccacct tgggaactga ccggttagtc tcgtcagcca gcattactgg
5761 cggccctgac aacagcacca ccctcactta tcgaaaggtt tcatgcttgc tgttaaagct
5821 gaatgtctct atgtgggatg agccaccgga actacagctg ctaggttccc agtctctccc
5881 taacattact aatattactc aaatttctgg tgtaaccgga ggatgcgtag gctttagacc
5941 aggagggatc ccctggtatc taggatggac tagacaggaa gccacacggt tcctccttag
6001 acaatcctcc ttttctaatt ccacggaacc atttacggtg gtgacagcgg atagacacaa
6061 tctgtttatg gggagtgagt actgtggtgc atatggctac agattttggg aagtatataa
6121 ctgctcacaa gaacagggtc agggccaaca gcaataccgc tgtggcaagg cacgcagccc
6181 ccgcccgggt tctcctgaaa cccagtgcat aaggagagga ggtaaatggg ttaatcaatc
6241 acaggagatt aatgaaacgg agccgttcag ctttacggtg aactgtacag ctagtaattt
6301 gcgtaatgcc agtgggtgct gcggaaaaac aggtatgatc ctcccagggg catggatcga
6361 cagcacaaaa ggtaatttca ccaaaccaaa agcgctacca cccggaattt tcctcatttg
6421 tggggatcgc gcatggcaag gaattcccag tcgtccggta gggggccccт gctatttagg
6481 caagcttacc atgttagcac ccaaccatac agatattctc aaaatccttg ttaattcatc
6541 gcgaacaggt attagacgta aacgaagcgt ctcacacctg gatgatacat gctcagatga
6601 ggtgcagctt tggggtccta cagcaagaat ctttgcatct atcttagccc cgggggtagc
6661 agctgcgcaa gccttaagag aaattgagag actagcctgt tggtccgtta aacaggctaa
6721 cttgacaaca tcactcctcg gggacttatt ggatgatgtc acgagtattc gacacgcggt
6781 cctgcagaac cgagcggcta ttgacttctt gctcctagct cacggccatg gctgtgagga
6841 cgttgccgga atgtgttgct tcaatctgag tgatcacagt gagtctatac agaagaagtt
6901 ccagcttatg aaggaacatg tcagtaagat cggcgtggac agcgacccaa tcggaagttg
6961 gctacgaggg ttattcgggg gaatagggga atgggccgtt catctgctaa aaggactgct
7021 tttggggctt gtagttattt tattgctagt agtgtgcctg ccttgccttt tacaatttgt
7081 gtctagtagc atccgaagga taattaatag ttcaatcagc tatcatactg aatacaggaa
7141 gatgcagggc ggagcagtct agagctcaat tataataatc ttgcgaatcg ggctgtaacg
7201 gggcaaggct tgaccgaggg gactataaca tgtataggcg aaaagcgggg tctcggttgt
7261 aacgcgctta ggaagtcccc tcgaggtatg gcagatatgc ttttgcatag ggaggggggaa
7321 atgtagtctt aatcgtaggt taacatgtat attaccaaat aagggaatcg cctgatgcac
7381 caaataaggt attatatgat cccattggtg gtgaaggagc gacctgaggg catatgggcg
7441 ttaacagaac tgtctgtcct tgcgtcattc ctcatcggat catgtacgcg gcagagtatg
7501 attggataac aggatggcac cattcatcgt ggcgcatgct gattggtgcg actaaggagt
7561 tgtgtaaccc acgaatgtac ttaagcttgt agttgctaac aataaagtgc cattctacct
7621 ctcaccacat tggtgtgcac ctgggttgat ggccggaccg tcgattccct gacgactgcg
7681 aacacctgaa tgaagctgaa ggcttcaa
//
```

Jul 8 2003 12:22:35

☐ **1:** L10923. Myeloblastosis-as...[gi:407979]

Links

```
LOCUS       MAVTYP1A2A              7704 bp    DNA     linear   VRL 18-APR-1994
DEFINITION  Myeloblastosis-associated virus type 1/2 env gene, complete cds;
            gag gene, complete cds; pol gene, 3' end of cds.
ACCESSION   L10923
VERSION     L10923.1  GI:407979
KEYWORDS    env gene; gag gene; long terminal repeat; pol gene.
SOURCE      Avian myeloblastosis-associated virus 1/2
  ORGANISM  Avian myeloblastosis-associated virus 1/2
            Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
REFERENCE   1  (bases 1 to 7704)
  AUTHORS   Joliot,V., Boroughs,K., Lasserre,F., Crochet,J., Dambrine,G.,
            Smith,R.E. and Perbal,B.
  TITLE     Pathogenic potential of myeloblastosis-associated virus:
            implication of env proteins for osteopetrosis induction
  JOURNAL   Virology 195 (2), 812-819 (1993)
  MEDLINE   93331743
   PUBMED   8393249
COMMENT     Original source text: Myeloblastosis-associated virus 1/2 (library:
            EMBL 4; MAV-2(0)/CEF) DNA.
FEATURES             Location/Qualifiers
     source          1..7704
                     /organism="Avian myeloblastosis-associated virus 1/2"
                     /proviral
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31668"
                     /cell_line="MAV-2(0) infected chicken embryo fibroblast"
                     /tissue_lib="EMBL 4; MAV-2(0)/CEF"
     LTR             1..390
     repeat_unit     1..292
                     /note="U3 repeat"
     repeat_unit     293..312
                     /note="R repeat"
     repeat_unit     313..390
                     /note="U5 repeat"
     gene            join(672..689,5370..7154)
                     /gene="env"
     CDS             join(672..689,5370..7154)
                     /gene="env"
                     /codon_start=1
                     /protein_id="AAA46300.1"
                     /db_xref="GI:407981"
                     /translation="MEAVIKAFLTGRPGKASKKDSKKKPPATSKKDPEKTPLLPTRVN
                     YILIIGVLVLCEVTGVRADVHLLEQPGNLWITWANRTGQTDFCLSTQSATSPFQTCLI
                     GIPSPISEGDFKGYVSDTNCTTLGTDRLVSSADFTGGPDNSTTLTYRKVSCLLLKLNV
                     SMWDEPPELQLLGSQSLPNITNIAQISGITGGCVGFRPQGVPWYLGWSRQEATRFLLR
                     HPSFSKSTEPFTVVTADRHNLFMGSEYCGAYGYRFWNMYNCSQVGRQYRCGNARSPRP
                     GLPEIQCTRRGGKWVNQSQEINESEPFSFTVNCTASSLGNAGGCCGKAGTILPGKWVD
                     STQGSFTKPKALPPGIFLICGDRAWQGIPSRPVGGPCYLGKLTMLAPNHTDILKILAN
```

```
                     SSRTGIRRKRSVSHLDDTCSDEVQLWGPTARIFASILAPGVAAAQALREIERLACWSV
                     KQANLTTSLLGDLLDDVASIRHAVLQNRAAIDFLLLAHGHGCEDVAGMCCFNLSDHSE
                     SIQKKFQLMKEHVNKIGVDSDPIGSWLRGLFGGIGEWAVHLLKGLLLGLVVILLLVVC
                     LPCLLQFVSSSIRRIINSSISYHTEYRKMQGGAV"
     gene            672..2777
                     /gene="gag"
     CDS             672..2777
                     /gene="gag"
                     /codon_start=1
                     /protein_id="AAA46299.1"
                     /db_xref="GI:407982"
                     /translation="MEAVIKVISSACKTYCGKTSPSKKEIGAMLSLLQKEGLLMSPSD
                     LYSPGSWDPITAALSQRAMVLGKSGELKTWGLVLGALKAAREEQVTSEQAKFWLGLGG
                     GRVSPPGPECIEKPATERRTDKGEEVGETTAHGDAKMAPEETATPKTVGTSCYHCGTA
                     IGCSCATASAPPPPYVGSGLYPSLAGVGEQQGQEGDTPRGAEQPRAEPGHAGLAPGPA
                     LTDWARVREELASTGPPVVAMPVVIKTEGPAWTPLEPKLITRLADTVRTKGLRSPITM
                     AEVEALMSSPLLPHDVTNLMRVILGPAPYALWMDAWGVQLQTVIAAATRDPRHPANGQ
                     GRGERTNLDRLKGLADGMVGNPQGQAALLRPGELVAITAAALQAFREVARLAEPAGPW
                     ADITQGPSEPFVDFANRLIKAVEGSDLPPSARAPVIIDCFRQKSQPDIQQLIRAAPST
                     LTTPGEIIKYVLDRQKTAPLTDQGIAAAMSSAIQPLVMAVVNRERDGQTGSGGRARGL
                     CYTCGSPGHYQAQCPKKRKSGNSRERCQLCDGMGHNAKQCRKRDGNQGQRPGRGLSSG
                     PWPVSEQPAVSLAMTMEHKDRPLVRVILTNTGSHPVKQRSVYITALLDSGADITIISE
                     EDWPADWPVMEAANPQIHGIGGGIPMRKSRDMIEVGVINRDGSLERPLLLFPAVAMVR
                     GSILGRDCLQGLGLRLTNL"
     exon            <672..689
                     /gene="gag"
     intron          690..5369
                     /gene="env"
     gene            2780..5482
                     /gene="pol"
     CDS             <2780..5482
                     /gene="pol"
                     /note="5' end undetermined"
                     /codon_start=1
                     /protein_id="AAA46301.1"
                     /db_xref="GI:407980"
                     /translation="RATVFTVALHLAIPLKWKPDHTPVWIDQWPLPEGKLVALTQLVE
                     KELQLGHIEPSLSCWNTPVFVIRKASGSYRLLHDLRAVNAKLVPFGAVQQGAPVLSAL
                     PRGWPLMVLDLKDCFFSIPLAEQDREAFAFTLPSVNNQAPARRFQWKVLPQGMTCSPT
                     ICQLIVGQILEPLRLKHPSLRMLHYMDDLLLAASSHDGLEAAGEEVISTLERAGFTIS
                     PDKVQKEPGVQYLGYKLGSTYVAPVGLVAEPRIATLWDVQKLVGSLQSVRPALGIPPR
                     LMGPFYEQLRGSDPNEAREWNLDMKMAWREIVQLSTTAALERWDPALPLEGAVARCEQ
                     GAIGVLGQGLSTHPRPCLWLFSTQPTKAFTAWLEVLTLLITKLRASAVRTFGKEVDIL
                     LLPACFREDLPLPEGILLALRGFAGKIRSSDTPSIFDIARPLHVSLKVRVTDHPVPGP
                     TVFTDASSSTHKGVVVWREGPRWEIKEIADLGASVQQLEARAVAMALLLWPTTPTNVV
                     TDSAFVAKMLLKMGQEGVPSTAAAFILEDALSQRSAMAAVLHVRSHSEVPGFFTEGND
                     VADSQATFQAYPLREAKDLHTALHIGPRALSKACNISMQQAREVVQTCPHCNSAPALE
                     AGVNPRGLGPLQIWQTDFTLEPRMAPRSWLAVTVDTASSAIVVTQHGRVTSVAAQHHW
                     ATAIAVLGRPKAIKTDNGSCFTSKSTREWLARWGIAHTTGIPGNSQGQAMVERANRLL
                     KDKIRVLAEGDGFMKRIPTSKQGELLAKAVYALNHFERGENTKTPIQKHWRPTVLTEG
                     PPVKIRIETGEWEKGWNVLVWGRGYAAVKNRDTDKVIWVPSRKVKPDITQKDEVTKRD
                     EASPLFAGISDWAPWEGEQEGLQEETASNKQERPGEDTLAANES"
     exon            5370..7704
                     /partial
BASE COUNT     1849 a   1862 c   2212 g   1781 t
ORIGIN
        1 tgtagtctta atcataggtt aacatgtata ttaccaaata agggaatcgc ctgatgcacc
       61 aaataaggta ttatatgatc ccattggtgg cgaaggagcg acctgagggc atatgggcgt
      121 taacagaact gtctgtcctt gcgtcattcc ttcctcatca gatcatgtac gcggcagagt
      181 atgattggat aactggatgg caccattcat cgtggcgcat gctgattggt gcgactaagg
      241 agttatgtaa cccacgaatg tacttaagct tgtagttgct aacaataaag tgccattcta
```

```
 301 cctctcacca cattggtgtg cacctgggtt gatggccgga ccgtcgattc cctgacgact
 361 acgagcacct gcatgaagca gaaggcttca tttggtgacc ccgacgtgat cgttagggaa
 421 tagtggtcgg ccacaggcgg cgtggcgatc ctgtctccat ccgtctcgtc gaccgggagg
 481 cgagttcgat gaccctgata gaggggggctg cggcttaggg aggcagaagc tgagcaccgt
 541 cggtaggagc tccagggccc ggagcggata cccctgccga gaactcagac agtcgtcgga
 601 agacgcggag ggagcccgac gattgagtgg tccgttgttc tgcgcgattc cggtcgcccg
 661 gtggatcaaa catggaagcc gtcataaagg tgatttcgtc cgcgtgtaaa acctattgcg
 721 ggaaaacctc tccttctaag aaggaaatag gggccatgtt gtccctgtta caaaaggaag
 781 ggttgcttat gtccccctca gacttatact ccccggggtc ctgggatcct attaccgcgg
 841 cgctctccca gcgagctatg gtacttggga aatcgggaga gttaaaaacc tggggattgg
 901 ttttgggggc attgaaggca gctcgagagg aacaggttac atctgagcaa gcgaagtttt
 961 ggttgggatt agggggagga agggtctctc ccccaggtcc ggagtgtatc gagaaaccag
1021 caacagagcg gcgaaccgac aaggggggagg aagtgggaga aacaactgcg cacggagatg
1081 cgaagatggc gccggaggaa acggccacac ctaagaccgt tggcacatcc tgctatcatt
1141 gcggaacagc tattggctgt agttgcgcca cagcctcggc tcctcctcct ccttatgtgg
1201 ggagtggttt gtatccttcc ctggcggggg tgggagagca gcagggccag gagggtgaca
1261 cacctcgggg ggcggaacag ccaagggcgg agccagggca cgcgggtctg gcccctgggc
1321 cggccctgac tgattgggca agggtcaggg aggagcttgc gagtactggt ccgcccgtgg
1381 tggccatgcc tgtagtgatt aagacagagg gacctgcctg gacccctctg gagccaaaat
1441 tgatcacaag actggctgat acggtcagga ccaagggctt acgatccccg atcactatgg
1501 cagaagttga agcgcttatg tcctccccgc tgctgccgca tgacgtcacg aatctaatga
1561 gagttatttt aggacctgcc ccatatgcct tatggatgga cgcttgggga gtccaattac
1621 agacggttat agcggcagcc actcgcgacc cccgacaccc agcgaacggt caagggcggg
1681 gggaacggac taacttggat cgcttaaagg gtctagctga tgggatggtg ggcaacccac
1741 agggtcaggc cgcattatta agaccggggg aattggttgc tattacggcg gcggctctcc
1801 aggcgtttag agaggttgcc cggctggcgg aacctgcagg tccgtgggcg gatatcacgc
1861 agggaccatc tgagcccttt gttgattttg ccaatcggct tataaaggcg gttgaggggt
1921 cagatctccc gccttccgcg cgggctccgg tgatcattga ctgctttagg cagaagtcac
1981 agccagatat ccagcagctt atacgggcag caccctccac gctgaccacc ccaggagaga
2041 taatcaaata tgtgctagac aggcagaaga ctgcccctct tacggatcaa ggcatagccg
2101 cggccatgtc gtctgctatc cagcccttag ttatggcagt agtcaataga gagagggatg
2161 gacaaactgg gtcgggtggt cgtgcccgag ggctctgcta cacttgtgga tccccgggac
2221 attatcaggc gcagtgcccg aaaaaacgga agtcaggaaa cagccgtgag cgatgtcagt
2281 tgtgtgacgg gatgggacac aacgctaaac agtgtaggaa gcgggatggc aaccagggcc
2341 aacgcccagg aagaggcctc tcttcggggc cgtggcccgt ctctgagcag cctgccgtct
2401 cgttagcgat gacaatggaa cataaagatc gccccttggt tagggtcatt ctgactaaca
2461 ctgggagtca tccagtcaaa caacgttcgg tgtatatcac cgcgctgttg gactctggag
2521 cggacatcac tattatttca gaggaggact ggcccgccga ttggccagtg atggaggctg
2581 cgaacccaca gatccatggg ataggagggg gaattcccat gcgaaaatcc cgggatatga
2641 tagaagtggg ggttattaac cgagacggat ccttggagcg acccctgctc ctcttcccgg
2701 cagtggctat ggttagaggg agtattctgg ggagagattg tctgcagggc ctagggctcc
2761 gcttgacaaa tttataggga gggccactgt tttcactgtt gcgctacatc tggctatccc
2821 gctcaaatgg aagccagacc acacgcctgt gtggattgac cagtggcccc tccctgaagg
2881 taaacttgta gcgctgacgc aattagtgga aaaagaatta cagttaggac atatagaacc
2941 ttcacttagt tgctggaaca cacctgtctt tgtgatccgg aaggcttccg ggtcttatcg
3001 cttattgcat gacttgcgcg ctgttaacgc taagcttgtt ccttttgggg ccgtccaaca
3061 gggggcgccg gttctctccg cgctcccgcg tggctggccc ctgatggtcc tagacctcaa
3121 ggattgcttc ttttctattc ctcttgcgga acaagatcgc gaagcttttg catttacgct
3181 cccctctgtg aataaccagg cccccgctcg aaggttccaa tggaaggtct tgccccaagg
3241 gatgacctgt tctcccacta tctgtcagtt gatagtgggt caaatacttg agcccttgcg
3301 actcaagcac ccatctctgc gcatgttgca ttatatggat gatcttttgc tagccgcctc
3361 aagtcatgat gggttggaag cggcagggga ggaggttatc agtacattgg aaagagccgg
3421 gttcaccatt tcgcctgata aggtccagaa ggagcccgga gtacaatatc ttgggtacaa
3481 gttaggtagt acgtatgtag cacccgtagg cctggtagca gaacccagga tagccacctt
3541 gtgggatgtt cagaagctgg tggggtcact tcagtccgtt cgcccagcgt taggaatccc
3601 gccacgactg atgggccccct tttatgagca gttacgaggg tcagatccta acgaggcgag
3661 ggaatggaat ctagacatga aaatggcctg gagagagatc gtacagctta gcaccactgc
3721 tgccttggag cgatgggacc ctgccttgcc tctggaagga gcggtcgcta gatgtgaaca
3781 gggggcaata ggggtcctgg gacagggact gtccacacac ccaaggccat gtttgtggct
3841 attctccacc caacccacca aggcgtttac tgcttggtta gaagtgctca ccctttgat
3901 tactaagcta cgtgcttcgg cagtgcgaac ctttggcaag gaggttgata tcctcctgtt
```

```
3961 gcctgcatgc tttcgggagg accttccgct cccggagggg atcctgttag cccttagggg
4021 gtttgcagga aaaatcagga gtagtgacac gccatctatt tttgacattg cgcgtccact
4081 gcatgtttct ctgaaagtga gggttaccga ccaccctgtg ccgggaccca ctgtctttac
4141 tgacgcctcc tcaagcaccc ataagggggt ggtagtctgg agggagggcc caaggtggga
4201 gataaaagaa atagctgatt tgggggcaag tgtacaacaa ctggaagcac gcgctgtggc
4261 catggcactt ctgctgtggc cgacaacgcc cactaatgta gtgactgact ctgcgtttgt
4321 tgcgaaaatg ttactcaaga tggggcagga gggagtcccg tctacagcgg cggcttttat
4381 tttagaggat gcgttaagcc aaaggtcagc catggccgcc gttctccacg tgcggagtca
4441 ttctgaagtg ccagggtttt tcacagaagg aaatgacgtg gcagatagcc aagccacctt
4501 tcaagcgtat cccttgagag aggctaaaga tctccatacc gctctccata ttggaccccg
4561 cgcgctatcc aaagcgtgta atatatctat gcagcaggct agggaggttg ttcagacctg
4621 cccgcattgt aattcagccc ctgcgttgga ggccggggta aaccctaggg gtttgggacc
4681 cctacagata tggcagacag actttacact tgagcctaga atggctcccc gttcctggct
4741 cgctgttact gtggataccg cctcatcggc gatagtcgta actcagcatg gccgtgtcac
4801 atcggttgct gcacaacatc attgggctac ggctatcgcc gttttgggaa gaccaaaggc
4861 cataaaaaca gataacgggt cctgcttcac gtctaaatcc acgcgagagt ggctcgcgag
4921 atgggggata gcacacacca ccgggattcc gggtaattcc cagggtcaag ctatggtaga
4981 gcgggccaac cggctcctga aagataagat ccgtgtgctt gcggaggggg atggctttat
5041 gaaaagaatc cccaccagca aacaggggga actattagcc aaggcagtgt atgcccttaa
5101 tcactttgag cgtggtgaaa acacaaaaac accgatacaa aaacattgga gacctaccgt
5161 tcttacagaa ggacccccgg ttaaaatacg aatagagaca ggggagtggg aaaaaggatg
5221 gaacgtgctg gtctggggac gaggttatgc cgctgtgaaa aacagggaca ctgataaggt
5281 tatttgggta ccctctcgaa aagttaaacc ggacatcacc caaaaggatg aggtgactaa
5341 gagagatgag gcaagccctc tttttgcagg catttctgac tgggcgccct gggaaggcga
5401 gcaagaagga ctccaagaag aaaccgccag caacaagcaa gaaagacccg gagaagacac
5461 ccttgctgcc aacgagagtt aattatattc tcattattgg tgtcttggtc ttgtgtgagg
5521 ttacgggggt aagagctgat gtccacttac tcgagcaacc agggaacctt tggattacat
5581 gggccaaccg tacaggccaa acggatttct gcctctctac acagtcggcc acctcccctt
5641 ttcaaacatg tttgataggt atcccgtccc ctatttccga gggtgatttt aagggatatg
5701 tttctgatac aaattgcacc accttgggaa ctgatcggtt agtctcgtca gccgacttta
5761 ctggcggacc tgacaacagt accaccctca cttatcggaa ggtctcatgc ttgttgttaa
5821 agctgaatgt ctctatgtgg gatgagccac ctgaactaca gctgttaggt tcccagtctc
5881 tccctaacat tactaatatt gctcagattt ccggtataac cgggggatgc gtaggcttca
5941 gaccacaagg ggttccttgg tatctaggtt ggtctagaca ggaggccacg cggtttctcc
6001 ttagacaccc ctctttctct aaatccacgg aaccgtttac ggtggtgaca gcggataggc
6061 acaatctttt tatggggagt gagtactgcg gtgcatatgg ctacagattt tggaacatgt
6121 ataactgctc acaggtgggg cggcagtacc gctgtggtaa tgcgcgcagc ccccgcccgg
6181 gtcttcctga aatccagtgt acaaggagag gaggcaaatg ggttaatcaa tcacaggaaa
6241 ttaatgagtc ggagccgttc agctttacgg tgaactgtac agctagtagt ttgggtaatg
6301 ccggtgggtg ttgcggaaaa gcaggcacga ttctcccggg aaagtgggtc gacagcacac
6361 aaggtagttt cactaaacct aaagcgctac cacccggaat tttcctcatt tgtggggatc
6421 gcgcatggca aggaattcct agtcgtccgg taggggccc ctgctattta ggcaagctta
6481 ccatgttagc acccaaccat acagatattc tcaaaatcct tgctaattca tcgcgaacag
6541 gtataagacg taaacgaagc gtctcacacc tggatgacac atgctcagat gaagtgcagc
6601 tttggggtcc tacagcaaga atctttgcat ctatcctagc cccgggggta gcagctgcgc
6661 aagccttaag agaaatcgag agactagcct gttggtccgt taaacaggct aatttgacaa
6721 catcactcct cggggactta ttggatgatg tcgcgagtat tcgacacgcg gtcctgcaga
6781 accgagcggc tattgacttc ttgcttctag ctcatggcca tggctgtgag gacgttgccg
6841 gaatgtgttg cttcaatctg agtgatcaca gtgagtctat acagaagaag ttccagctta
6901 tgaaggaaca tgtcaataag atcggcgtgg acagcgaccc aatcggaagt tggctacgag
6961 ggttattcgg gggaataggg gaatgggccg ttcatctgct aaaaggactg cttttggggc
7021 ttgtagttat tttattgcta gtagtgtgcc tgccttgcct tttgcaattt gtgtctagta
7081 gcatccgaag gataattaat agttcaatca gctatcatac tgaatacagg aagatgcagg
7141 gcggagcagt ctagagctca attataataa tcttgcgaat cgggctgtaa cggggcaagg
7201 cttgaccgag gggactataa catgtatagg cgaaaagcgg ggtctcggtt gtaacgcgct
7261 taggaagtcc cctcgaggta tggcagacat gcttttgcat agggaggggg aaatgtagtc
7321 ttaatcatag gttaacatgt atattaccaa ataagggaat cgcctgatgc accaaataag
7381 gtattatatg atcccattgg tggcgaagga gcgacctgag ggcatatggg cgttaacaga
7441 actgtctgtc cttgcgtcat tccttcctca tcagatcatg tacgcggcag agtatgattg
7501 gataactgga tggcaccatt catcgtggcg catgctgatt ggtgcgacta aggagttatg
7561 taacccacga atgtacttaa gcttgtagtt gctaacaata aagtgccatt ctacctctca
```

```
     7621 ccacattggt gtgcacctgg gttgatggcc ggaccgtcga ttccctgacg actacgagca
     7681 cctgcatgaa gcagaaggct tcaa
//
```

☐ **1:** L10924. Myeloblastosis-as...[gi:407987]

Links

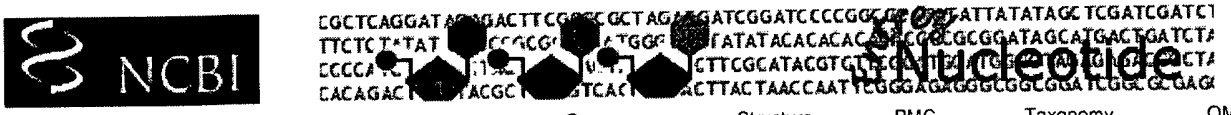
```
LOCUS       MAVTYP2AAA               7726 bp    DNA     linear   VRL 18-APR-1994
DEFINITION  Myeloblastosis-associated virus type 2 env gene, complete cds; gag
            gene, complete cds; pol gene, 3' end of cds.
ACCESSION   L10924
VERSION     L10924.1  GI:407987
KEYWORDS    env gene; gag gene; long terminal repeat; pol gene.
SOURCE      Avian myeloblastosis-associated virus type 2
  ORGANISM  Avian myeloblastosis-associated virus type 2
            Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
REFERENCE   1  (bases 1 to 7726)
  AUTHORS   Joliot,V., Boroughs,K., Lasserre,F., Crochet,J., Dambrine,G.,
            Smith,R.E. and Perbal,B.
  TITLE     Pathogenic potential of myeloblastosis-associated virus:
            implication of env proteins for osteopetrosis induction
  JOURNAL   Virology 195 (2), 812-819 (1993)
  MEDLINE   93331743
   PUBMED   8393249
COMMENT     Original source text: Myeloblastosis-associated virus type 2 DNA.
FEATURES             Location/Qualifiers
     source          1..7726
                     /organism="Avian myeloblastosis-associated virus type 2"
                     /proviral
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31670"
                     /cell_line="MAV-2 (0) infected chicken embryo fibroblast"
     LTR             1..390
     repeat_unit     1..292
                     /note="U3 repeat"
     repeat_unit     293..312
                     /note="R repeat"
     repeat_unit     313..390
                     /note="U5 repeat"
     exon            <672..690
     gene            join(673..690,5371..7176)
                     /gene="env"
     CDS             join(673..690,5371..7176)
                     /gene="env"
                     /codon_start=1
                     /protein_id="AAA46306.1"
                     /db_xref="GI:407989"
                     /translation="MEAVIKAFLTGRPGKASKKDSKKKPPATSKKDPEKTPLLPTRVN
                     YILIIGVLVLCEVTGVRADVHLLEQPGNLWITWANRTGQTDFCLSTQSATSPFQTCLI
                     GIPSPISEGDFKGYVSDTDCTTLGTHRLVSRGIPGGPDNSTTLTYRKVSCLLLKLNVS
                     LLDEPSELQLLGSQSLPNITDITRIPTVAGGCIGFTPYGSPAGVYGWDRRQVTHILLT
                     DPGNNPFFDKASNSSKPFTVVTADRHNLFMGSEYCGAYGYRFWEMYNCSQMSQNWSIC
                     QDVWGRGLPESWCTSTGGIWVNQSKAINETEPFSFTANCTGSNLGNVSGCCGESITLL
                     PPGAWVDSTQGSFTKPKALPPGIFLICGDRAWQGIPSRPVGGPCYLGKLTMLAPNHTD
                     ILKILANSSRTGIRRKRSVSHLDDTCSDEVQLWGPTARIFASILAPGVAAAQALREIE
```

```
                     RLACWSVKQANLTTSLLGDLLDDVASIRHAVLQNRAAIDFLLLAHGHGCEDVAGMCCF
                     NLSDHSESIQKKFQLMKEHVNKIGVDSDPIGSWLRGLFGGIGEWAVHLLKGLLLGLVV
                     ILLLVVCLPCLLQFVSSSIRRIINSSISYHTEYRKMQGGAV"
     gene            673..2778
                     /gene="gag"
     CDS             673..2778
                     /gene="gag"
                     /codon_start=1
                     /protein_id="AAA46305.1"
                     /db_xref="GI:407990"
                     /translation="MEAVIKVISSACKTYCGKTSPSKKEIGAMLSLLQKEGLLMSPSD
                     LYSPGSWDPITAALSQRAMVLGKSGELKTWGLVLGALKAAREEQVTSEQAKFWLGLGG
                     GRVSPPGPECIEKPATERRTDKGEEVGETTAHGDAKMAPEETATPKTVGTSCYHCGTA
                     IGCSCATASAPPPPYVGSGLYPSLAGVGEQQGQEGDTPRGAEQPRAEPGHAGLAPGPA
                     LTDWARVREELASTGPPVVAMPVVIKTEGPAWTPLEPKLITRLADTVRTKGLRSPITM
                     AEVEALMSSPLLPHDVTNLMRVILGPAPYALWMDAWGVQLQTVIAAATRDPRHPANGQ
                     GRGERTNLDRLKGLADGMVGNPQGQAALLRPGELVAITAAALQAFREVARLAEPAGPW
                     ADITQGPSEPFVDFANRLIKAVEGSDLPPSARAPVIIDCFRQKSQPDIQQLIRAAPST
                     LTTPGEIIKYVLDRQKTAPLTDQGIAAAMSSAIQPLVMAVVNRERDGQTGSGGRARGL
                     CYTCGSPGHYQAQCPKKRKSGNSRERCQLCDGMGHNAKQCRKRDGNQGQRPGRGLSSG
                     PWPVSEQPAVSLAMTMEHKDRPLVRVILTNTGSHPVKQRSVYITALLDSGADITIISE
                     EDWPADWPVMEAANPQIHGIGGGIPMRKSRDMIEVGVINRDGSLERPLLLFPAVAMVR
                     GSILGRDCLQGLGLRLTNL"
     intron          691..5368
                     /gene="env"
     gene            2776..5483
                     /gene="pol"
     CDS             <2776..5483
                     /gene="pol"
                     /note="5' end undetermined"
                     /codon_start=3
                     /protein_id="AAA46307.1"
                     /db_xref="GI:407988"
                     /translation="GRATVFTVALHLAIPLKWKPDHTPVWIDQWPLPEGKLVALTQLV
                     EKELQLGHIEPSLSCWNTPVFVIRKASGSYRLLHDLRAVNAKLVPFGAVQQGAPVLSA
                     LPRGWPLMVLDLKDCFFSIPLAEQDREAFAFTLPSVNNQAPARRFQWKVLPQGMTCSP
                     TICQLIVGQILEPLRLKHPSLRMLHYMDDLLLAASSHDGLEAAGEEVISTLERAGFTI
                     SPDKVQKEPGVQYLGYKLGSTYVAPVGLVAEPRIATLWDVQKLVGSLQSVRPALGIPP
                     RLMGPFYEQLRGSDPNEAREWNLDMKMAWREIVQLSTTAALERWDPALPLEGAVARCE
                     QGAIGVLGQGLSTHPRPCLWLFSTQPTKAFTAWLEVLTLLITKLRASAVRTFGKEVDI
                     LLLPACFREDLPLPEGILLALRGFAGKIRSSDTPSIFDIARPLHVSLKVRVTDHPVPG
                     PTVFTDASSSTHKGVVVWREGPRWEIKEIADLGASVQQLEARAVAMALLLWPTTPTNV
                     VTDSAFVAKMLLKMGQEGVPSTAAAFILEDALSQRSAMAAVLHVRSHSEVPGFFTEGN
                     DVADSQATFQAYPLREAKDLHTALHIGPRALSKACNISMQQAREVVQTCPHCNSAPAL
                     EAGVNPRGLGPLQIWQTDFTLEPRMAPRSWLAVTVDTASSAIVVTQHGRVTSVAAQHH
                     WATAIAVLGRPKAIKTDNGSCFTSKSTREWLARWGIAHTTGIPGNSQGQAMVERANRL
                     LKDKIRVLAEGDGFMKRIPTSKQGELLAKAVYALNHFERGENTKTPIQKHWRPTVLTE
                     GPPVKIRIETGEWEKGWNVLVWGRGYAAVKNRDTDKVIWVPSRKVKPDITQKDEVTKR
                     DEASPLFAGISDWAPWEGEQEGLQEETASNKQERPGEDTLAANES"
     exon            5369..>7726
BASE COUNT     1864 a    1868 c    2206 g    1788 t
ORIGIN
        1 tgtagtctta atcataggtt aacatgtata ttaccaaata agggaatcgc ctgatgcacc
       61 aaataaggta ttatatgatc ccattggtgg cgaaggagcg acctgagggc atatgggcgt
      121 taacagaact gtctgtcctt gcgtcattcc ttcctcatca gatcatgtac gcggcagagt
      181 atgattggat aactggatgg caccattcat cgtggcgcat gctgattggt gcgactaagg
      241 agttatgtaa cccacgaatg tacttaagct tgtagttgct aacaataaag tgccattcta
      301 cctctcacca cattggtgtg cacctgggtt gatggccgga ccgtcgattc cctgacgact
      361 acgagcacct gcatgaagca gaaggcttca atttggtgac cccgacgtga tcgttaggga
      421 atagtggtcg gccacaggcg gcgtggcgat cctgtctcca tccgtctcgt cgaccgggag
      481 gcgagttcga tgaccctgat agaggggct gcggcttagg gaggcagaag ctgagcaccg
```

```
 541 tcggtaggag ctccagggcc cggagcggat acccctgccg agaactcaga cagtcgtcgg
 601 aagacgcgga gggagcccga cgattgagtg gtccgttgtt ctgcgcgatt ccggtcgccc
 661 ggtggatcaa acatggaagc cgtcataaag gtgatttcgt ccgcgtgtaa aacctattgc
 721 gggaaaacct ctccttctaa gaaggaaata ggggccatgt tgtccttgtt acaaaaggaa
 781 gggttgctta tgtccccctc agacttatac tccccggggt cctgggatcc tattaccgcg
 841 gcgctctccc agcgagctat ggtacttggg aaatcgggag agttaaaaac ctggggattg
 901 gttttggggg cattgaaggc agctcgagag gaacaggtta catctgagca agcgaagttt
 961 tggttgggat taggggggagg aagggtctct cccccaggtc cggagtgtat cgagaaacca
1021 gcaacagagc ggcgaaccga caaggggggag gaagtgggag aaacaactgc gcacggagat
1081 gcgaagatgg cgccggagga aacggccaca cctaagaccg ttggcacatc ctgctatcat
1141 tgcggaacag ctattggctg tagttgcgcc acagcctcgg ctcctcctcc tccttatgtg
1201 gggagtggtt tgtatccttc cctggcgggg gtgggagagc agcagggcca ggagggtgac
1261 acacctcggg gggcggaaca gccaagggcg gagccagggc acgcgggtct ggcccctggg
1321 ccggccctga ctgattgggc aagggtcagg gaggagcttg cgagtactgg tccgcccgtg
1381 gtggccatgc ctgtagtgat taagacagag ggacctgcct ggacccctct ggagccaaaa
1441 ttgatcacaa gactggctga tacggtcagg accaagggct tacgatcccc gatcactatg
1501 gcagaagttg aagcgcttat gtcctccccg ctgctgccgc atgacgtcac gaatctaatg
1561 agagttattt taggacctgc cccatatgcc ttatggatgg acgcttgggg agtccaatta
1621 cagacggtta tagcggcagc cactcgcgac ccccgacacc cagcgaacgg tcaagggcgg
1681 ggggaacgga ctaacttgga tcgcttaaag ggtctagctg atgggatggt gggcaaccca
1741 cagggtcagg ccgcattatt aagaccgggg gaattggttg ctattacggc ggcggctctc
1801 caggcgttta gagaggttgc ccggctggcg gaacctgcag gtccgtgggc ggatatcacg
1861 cagggaccat ctgagccctt tgttgatttt gccaatcggc ttataaaggc ggttgagggg
1921 tcagatctcc cgccttccgc gcgggctccg gtgatcattg actgctttag gcagaagtca
1981 cagccagata tccagcagct tatacgggca gcaccctcca cgctgaccac cccaggagag
2041 ataatcaaat atgtgctaga caggcagaag actgcccctc ttacggatca aggcatagcc
2101 gcggccatgt cgtctgctat ccagccctta gttatggcag tagtcaatag agagagggat
2161 ggacaaactg ggtcgggtgg tcgtgcccga gggctctgct acacttgtgg atccccggga
2221 cattatcagg cgcagtgccc gaaaaaacgg aagtcaggaa acagccgtga gcgatgtcag
2281 ttgtgtgacg ggatgggaca caacgctaaa cagtgtagga agcgggatgg caaccagggc
2341 caacgcccag gaagaggcct ctcttcgggg ccgtggcccg tctctgagca gcctgccgtc
2401 tcgttagcga tgacaatgga acataaagat cgccccttgg ttagggtcat tctgactaac
2461 actgggagtc atccagtcaa acaacgttcg gtgtatatca ccgcgctgtt ggactctgga
2521 gcggacatca ctattatttc agaggaggac tggcccgccg attggccagt gatggaggct
2581 gcgaacccac agatccatgg gataggaggg ggaattccca tgcgaaaatc ccgggatatg
2641 atagaagtgg gggttattaa ccgagacgga tccttggagc gacccctgct cctcttcccc
2701 gcagtggcta tggttagagg gagtattctg gggagagatt gtctgcaggg cctagggctc
2761 cgcttgacaa atttataggg agggccactg ttttcactgt tgcgctacat ctggctatcc
2821 cgctcaaatg gaagccagac cacacgcctg tgtggattga ccagtggccc ctccctgaag
2881 gtaaacttgt agcgctgacg caattagtgg aaaaagaatt acagttagga catatagaac
2941 cttcacttag ttgctggaac acacctgtct ttgtgatccg gaaggcttcc gggtcttatc
3001 gcttattgca tgacttgcgc gctgttaacg ctaagcttgt tccttttggg gccgtccaac
3061 agggggcgcc ggttctctcc gcgctcccgc gtggctggcc cctgatggtc ctagacctca
3121 aggattgctt cttttctatt cctcttgcgg aacaagatcg cgaagctttt gcatttacgc
3181 tcccctctgt gaataaccag gcccccgctc gaaggttcca atggaaggtc ttgccccaag
3241 ggatgacctg ttctcccact atctgtcagt tgatagtggg tcaaatactt gagcccttgc
3301 gactcaagca cccatctctg cgcatgttgc attatatgga tgatcttttg ctagccgcct
3361 caagtcatga tgggttggaa gcggcagggg aggaggttat cagtacattg gaaagagccg
3421 ggttcaccat ttcgcctgat aaggtccaga aggagcccgg agtacaatat cttgggtaca
3481 agttaggtag tacgtatgta gcacccgtag gcctggtagc agaacccagg atagccacct
3541 tgtgggatgt tcagaagctg gtggggtcac ttcagtccgt tcgcccagcg ttaggaatcc
3601 cgccacgact gatgggcccc ttttatgagc agttacgagg gtcagatcct aacgaggcga
3661 gggaatggaa tctagacatg aaaatggcct ggagagagat cgtacagctt agcaccactg
3721 ctgccttgga gcgatgggac cctgccttgc ctctggaagg agcggtcgct agatgtgaac
3781 agggggcaat aggggtcctg ggacagggac tgtccacaca cccaaggcca tgtttgtggc
3841 tattctccac ccaacccacc aaggcgttta ctgcttggtt agaagtgctc acccttttga
3901 ttactaagct acgtgcttcg gcagtgcgaa cctttggcaa ggaggttgat atcctcctgt
3961 tgcctgcatg ctttcgggag gaccttccgc tcccggaggg gatcctgtta gcccttaggg
4021 ggtttgcagg aaaaatcagg agtagtgaca cgccatctat ttttgacatt gcgcgtccac
4081 tgcatgtttc tctgaaagtg agggttaccg accaccctgt gccgggaccc actgtcttta
4141 ctgacgcctc ctcaagcacc cataaggggg tggtagtctg gagggagggc ccaaggtggg
```

```
4201 agataaaaga aatagctgat ttgggggcaa gtgtacaaca actggaagca cgcgctgtgg
4261 ccatggcact tctgctgtgg ccgacaacgc ccactaatgt agtgactgac tctgcgtttg
4321 ttgcgaaaat gttactcaag atggggcagg agggagtccc gtctacagcg gcggctttta
4381 ttttagagga tgcgttaagc caaaggtcag ccatggccgc cgttctccac gtgcggagtc
4441 attctgaagt gccagggttt ttcacagaag gaaatgacgt ggcagatagc caagccacct
4501 ttcaagcgta tcccttgaga gaggctaaag atctccatac cgctctccat attggacccc
4561 gcgcgctatc caaagcgtgt aatatatcta tgcagcaggc tagggaggtt gttcagacct
4621 gcccgcattg taattcagcc cctgcgttgg aggccggggt aaaccctagg ggtttgggac
4681 ccctacagat atggcagaca gactttacac ttgagcctag aatggctccc cgttcctggc
4741 tcgctgttac tgtggatacc gcctcatcgg cgatagtcgt aactcagcat ggccgtgtca
4801 catcggttgc tgcacaacat cattgggcta cggctatcgc cgttttggga agaccaaagg
4861 ccataaaaac agataacggg tcctgcttca cgtctaaatc cacgcgagag tggctcgcga
4921 gatgggggat agcacacacc accgggattc cgggtaattc ccagggtcaa gctatggtag
4981 agcgggccaa ccggctcctg aaagataaga tccgtgtgct tgcggagggg gatggcttta
5041 tgaaaagaat ccccaccagc aaacaggggg aactattagc caaggcagtg tatgccctta
5101 atcactttga gcgtggtgaa aacacaaaaa caccgataca aaaacattgg agacctaccg
5161 ttcttacaga aggacccccg gttaaaatac gaatagagac aggggagtgg gaaaaaggat
5221 ggaacgtgct ggtctgggga cgaggttatg ccgctgtgaa aaacagggac actgataagg
5281 ttatttgggt accctctcga aaagttaaac cggacatcac ccaaaaggat gaggtgacta
5341 agagagatga ggcaagccct ctttttgcag gcatttctga ctgggcgccc tgggaaggcg
5401 agcaagaagg actccaagaa gaaaccgcca gcaacaagca agaaagaccc ggagaagaca
5461 cccttgctgc caacgagagt taattatatt ctcattattg gtgtcttggt cttgtgtgag
5521 gttacggggg taagagctga tgtccactta ctcgagcaac cagggaacct ttggattaca
5581 tgggccaacc gtacaggcca aacggatttc tgcctctcta cacagtcggc cacctcccct
5641 tttcaaacat gtttgatagg tatcccgtcc cctatttccg aaggtgactt taagggatat
5701 gtctctgata cagattgcac cactttggga actcaccggt tagtctcgag aggtattccg
5761 ggcggccctg ataacagcac aactctcacc tatcggaagg tttcatgctt attgttaaag
5821 ctgaacgttt ctctgttgga cgagccatca gaactacaac tgctgggttc ccagtctctc
5881 cccaacataa ctgatattac tcggatccct actgtggctg ggggatgcat aggctttacc
5941 ccatacggta gtccggctgg tgtttacgga tgggaccgga gacaggttac acacatcctt
6001 ttgaccgacc cagggaacaa tcctttcttt gataaggcct ctaactcctc gaaaccgttt
6061 accgtagtga cagcggacag gcacaatctc tttatgggga gtgagtattg tggtgcatat
6121 ggctacaggt tctgggaaat gtacaattgc tcacaaatga gtcagaattg gtccatctgt
6181 caggatgtgt ggggccgagg cctccctgaa agttggtgta caagcacagg aggtatatgg
6241 gttaatcaat cgaaggcaat caacgagaca gaaccgttca gttttactgc gaactgtact
6301 ggcagtaatt tgggtaatgt cagcggatgt tgcggagaat cgatcacact tctcccacca
6361 ggggcgtggg tcgacagcac acaaggtagt ttcactaaac ctaaagcgct accacccgga
6421 attttcctca tttgtgggga tcgcgcatgg caaggaattc ctagtcgtcc ggtagggggc
6481 ccctgctatt taggcaagct taccatgtta gcacccaacc atacagatat tctcaaaatc
6541 cttgctaatt catcgcgaac aggtataaga cgtaaacgaa gcgtctcaca cctggatgat
6601 acatgctcag atgaagtgca gctttggggt cctacagcaa gaatctttgc atctatccta
6661 gccccggggg tagcagctgc gcaagcctta agagaaatcg agagactagc ctgttggtcc
6721 gttaaacagg ctaatttgac aacatcactc ctcggggact tattggatga tgtcgcgagt
6781 attcgacacg cggtcctgca gaaccgagcg gctattgact tcttgcttct agctcatggc
6841 catggctgtg aggacgttgc cggaatgtgt tgcttcaatc tgagtgatca cagtgagtct
6901 atacagaaga agttccagct tatgaaggaa catgtcaata agatcggcgt ggacagcgac
6961 ccaatcggaa gttggctacg agggttattc gggggaatag gggaatgggc cgttcatctg
7021 ctaaaaggac tgcttttggg gcttgtagtt attttattgc tagtagtgtg cctgccttgc
7081 cttttacaat ttgtgtctag tagcatccga aggataatta atagttcaat cagctatcat
7141 actgaataca ggaagatgca gggcggagca gtctagagct caattataat aatcttgcga
7201 atcgggctgt aacggggcaa ggcttgaccg aggggactat aacatgtata ggcgaaaagc
7261 ggggtctcgg ttgtaacgcg cttaggaagt cccctcgagg tatggcagat atgcttttgc
7321 atagggaggg ggaaatgtag tcttaatcat aggttaacat gtatattacc aaataaggga
7381 atcgcctgat gcaccaaata aggtattata tgatcccatt ggtggcgaag gagcgacctg
7441 agggcatatg ggcgttaaca gaactgtctg tccttgcgtc attccttcct catcagatca
7501 tgtacgcggc agagtatgat tggataactg gatggcacca ttcatcgtgg cgcatgctga
7561 ttggtgcgac taaggagtta tgtaacccac gaatgtactt aagcttgtag ttgctaacaa
7621 taaagtgcca ttctacctct caccacattg gtgtgcacct gggttgatgg ccggaccgtc
7681 gattccctga cgactacgag cacctgcatg aagcagaagg cttcaa
//
```

Jul 8 2003 12:22:35



Links

☐ **1:** J02342. Rous sarcoma viru...[gi:210171]

```
LOCUS       ALRCG                   9625 bp ss-RNA     linear   VRL 04-MAR-1996
DEFINITION  Rous sarcoma virus (Prague strain, subgroup C) cDNA to genomic RNA,
            complete genome.
ACCESSION   J02342 J02021 J02343
VERSION     J02342.1  GI:210171
KEYWORDS    c-myc proto-oncogene; complete genome; env protein; gag protein;
            long terminal repeat; origin of replication; pol protein;
            polyprotein; protein kinase; src oncogene.
SOURCE      Rous sarcoma virus
  ORGANISM  Rous sarcoma virus
            Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
REFERENCE   1  (bases 234 to 351)
  AUTHORS   Haseltine,W.A., Maxam,A.M. and Gilbert,W.
  TITLE     Rous sarcoma virus genome is terminally redundant: the 5' sequence
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 74 (3), 989-993 (1977)
  MEDLINE   77149036
   PUBMED   66683
REFERENCE   2  (bases 234 to 351)
  AUTHORS   Shine,J., Czernilofsky,A.P., Friedrich,R., Bishop,J.M. and
            Goodman,H.M.
  TITLE     Nucleotide sequence at the 5' terminus of the avian sarcoma virus
            genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 74 (4), 1473-1477 (1977)
  MEDLINE   77172915
   PUBMED   67601
REFERENCE   3  (bases 1 to 374; 9273 to 9625)
  AUTHORS   Katz,R.A., Omer,C.A., Weis,J.H., Mitsialis,S.A., Faras,A.J. and
            Guntaka,R.V.
  TITLE     Restriction endonuclease and nucleotide sequence analyses of
            molecularly cloned unintegrated avian tumor virus DNA: structure of
            large terminal repeats in circle junctions
  JOURNAL   J. Virol. 42 (1), 346-351 (1982)
  MEDLINE   82217008
   PUBMED   6283156
REFERENCE   4  (bases 234 to 9545)
  AUTHORS   Schwartz,D.E., Tizard,R. and Gilbert,W.
  TITLE     Nucleotide sequence of Rous sarcoma virus
  JOURNAL   Cell 32 (3), 853-869 (1983)
  MEDLINE   83155662
   PUBMED   6299578
REFERENCE   5  (sites)
  AUTHORS   Broome,S. and Gilbert,W.
  TITLE     Rous sarcoma virus encodes a transcriptional activator
  JOURNAL   Cell 40 (3), 537-546 (1985)
  MEDLINE   85124605
   PUBMED   2982497
COMMENT     Original source text: Rous sarcoma virus (strain Prague) (clone:
            pATV-[6,8,9]) cDNA to genomic RNA.
```

[5] sites; transcriptional activator protein and mRNA. Proviral RSV has the following structure:
5'LTR-gag-pol-env-src-3'LTR. The single plus stranded 35S virion RNA is identical, with the exception of lacking the 5' U-3 and 3' U-5 segments. Two identical copies of this 35S RNA, associated as a 70S RNA complex, are present in each virion. After viral infection, the 35S RNA is reverse transcribed in the cytoplasm into linear double stranded DNA with a complete LTR at each end. After migrating to the nucleus, some of these molecules become circularized. The double stranded DNA integrates into the host DNA by an unknown mechanism.
Positions 335-352 are complementary to the 3' stem of host-encoded Trp-tRNA. Trp-tRNA binds to virion RNA at this site and serves as a primer for DNA synthesis by reverse transcription. The integrated proviral DNA is transcribed to produce 35S RNAs with sequence identical to the virion RNA. The 35S RNAs can be translated directly or processed by the cellular RNA splicing machinery to produce mRNAs encoding additional viral proteins. In order of length the viral mRNAs (and their products) are: 35Sa (gag-Pr76) mRNA, 35Sb (gag-pol-Pr180) mRNA, 35Sc (trn-act) mRNA, 28S (env-Pr95) mRNA, and 21S (src-p60) mRNA. The 35Sa (gag-Pr76) mRNA is apparently full-length and identical to virion RNA. The 35Sc (trn-act), 28S (env-Pr95), and 21S (src-p60) mRNAs all begin with the same 5' exon, but have varying lengths of intronic sequence (beginning at the splice-donor site following position 630) removed to produce the mature mRNAs.
The mechanism for production of the polyprotein precursor gag-pol-Pr-180 remains uncertain. The reading frame of gag is not the same as that of pol, so merely suppressing the amber stop codon of gag does not give a gag-pol read-through product, and no acceptable RNA splicing sites are apparent. It is known that the gag proteins, including P12, are cleaved from the gag-pol-Pr180 polyprotein as well as from the gag-Pr76 polyprotein. gag-pol-Pr180 is tentatively annotated in the features as two exons with undetermined intron boundaries. The pol-derived portion of gag-pol-Pr180 is processed to yield the reverse transcriptase beta subunit, which in turn is processed to yield the reverse transcriptase alpha subunit and p32 (tentatively identified as a DNA endonuclease).
[4] reports the dimer linkage site to be at position 756-781. The src gene is believed to have been obtained from avian DNA when an ALV-like virus recombined with host DNA. Homology to the c-src gene of chicken begins at position 7271. A direct repeat of about 100 bp is present near either end of exon 2 of the 21S (src) mRNA. The repeats include positions 7130-7222 and 9024-9123. [4] also sequenced 95% of the genome of Prague C RSV using cDNA to viral RNA. There are numerous conflicts between the sequence obtained from DNA and that obtained from cDNA. The sequence reported here is the DNA sequence.
[4] contains an in-depth discussion of the proteins encoded by gag, gag-pol, env, and src mRNAs. [4] points out 12 p19 binding sites that may influence splicing, RNA packing, and 35S RNA dimer formation.
[5] found that the target for the action of the transcriptional activator protein lies between 111 and 620 nucleotides upstream of the cap site.
Complete source information:
Rous sarcoma virus (Prague subgroup C): cDNA to viral RNA [2],[1], [4]; unintegrated DNA, clones pATV-6 [3], pATV-8 [3],[4],[5], and pATV-9 [3].

FEATURES Location/Qualifiers

```
source          1..9625
                /organism="Rous sarcoma virus"
                /mol_type="genomic RNA"
                /strain="Prague"
                /db_xref="taxon:11886"
                /clone="pATV-[6,8,9]"
LTR             1..334
                /note="5' LTR"
mRNA            234..9545
                /product="35Sa (gag), 35Sb (gag-pol), 35Sc (trn-act), 28S
                (env), 21S (src) mRNA"
repeat_region   234..254
                /note="5' terminal repeat"
conflict        279
                /citation=[4]
                /replace=""
misc_binding    335..352
                /bound_moiety="Trp-tRNA stem"
conflict        335
                /citation=[2]
                /replace=""
CDS             join(613..630,5311..7098)
                /note="env-Pr95 polyprotein precursor"
                /codon_start=1
                /protein_id="AAB59934.1"
                /db_xref="GI:210174"
                /translation="MEAVIKAFLTGYPGKTSKKDSKEKPLATSKKDPEKTPLLPTRVN
                YILIIGVLVLCEVTGVRADVHLLEQPGNLWITWANRTGQTDFCLSTQSATSPFQTCLI
                GIPSPISEGDFKGYVSDTNCSTVGTDRLVLSASITGGPDNSTTLTYRKVSCLLLKLNV
                SMWDEPPELQLLGSQSLPNVTNITQVSGVAGGCVYFAPRATGLFLGWSKQGLSRFLLR
                HPFTSTSNSTEPFTVVTADRHNLFMGSEYCGAYGYRFWEIYNCSQTRNTYRCGDVGGT
                GLPETWCRGKGGIWVNQSKEINETEPFSFTANCTGSNLGNVSGCCGEPITILPLGAWI
                DSTQGSFTKPKALPPAIFLICGDRAWQGIPSRPVGGPCYLGKLTMLAPNHTDILKILA
                NSSRTGIRRKRSVSHLDDTCSDEVQLWGPTARIFASILAPGVAAAQALREIERLACWS
                VKQANLTTSLLGDLLDDVTSIRHAVLQNRAAIDFLLLAHGHGCEDVAGMCCFNLSDHS
                ESIQKKFQLMKKHVNKIGVDSDPIGSWLRGIFGGIGEWAVHLLKGLLLGLVVILLLLV
                CLPCLLQFVSSSIRKMINSSINYHTEYRKMQGGAV"
mat_peptide     5479..7095
                /product="glycoprotein-85"
mat_peptide     6502..7095
                /product="glycoprotein-37"
CDS             join(613..2343,2736..5423)
                /note="gag-pol-Pr180 polyprotein precursor"
                /codon_start=1
                /protein_id="AAB59933.1"
                /db_xref="GI:210173"
                /translation="MEAVIKVISSACKTYCGKTSPSKKEIGAMLSLLQKEGLLMSPSD
                LYSPGSWDPITAALSQRAMILGKSGELKTWGLVLGALKAAREEQVTSEQAKFWLGLGG
                GRVSPPGPECIEKPATERRIDKGEEVGETTVQRDAKMAPEETATPKTVGTSCYHCGTA
                IGCNCATASAPPPPYVGSGLYPSLAGVGEQQGQGGDTPPGAEQSRAEPGHAGQAPGPA
                LTDWARVREELASTGPPVVAMPVVIKTEGPAWTPLEPKLITRLADTVRTKGLRSPITM
                AEVEALMSSPLLPHDVTNLMRVILGPAPYALWMDAWGVQLQTVIAAATRDPRHPANGQ
                GRGERTNLNRLKGLADGMVGNPQGQAALLRPGELVAITASALQAFREVARLAEPAGPW
                ADIMQGPSESFVDFANRLIKAVEGSDLPPSARAPVIIDCFRQKSQPDIQQLIRTAPST
                LTTPGEIIKYVLDRQKTAPLTDQGIAAAMSSAIQPLIMAVVNRERDGQTGSGGRARGL
                CYTCGSPGHYQAQCPKKRKSGNSRERCQLCNGMGHNAKQCRKRDGNQGQRPGKGLSSG
                PWPGPEPPAVSTVALHLAIPLKWKPDHTPVWIDQWPLPEGKLVALTQLVEKELQLGHI
                EPSLSCWNTPVFVIRKASGSYRLLHDLRAVNAKLVPFGAVQQGAPVLSALPRGWPLMV
                LDLKDCFFSIPLAEQDREAFAFTLPSVNNQAPARRFQWKVLPQGMTCSPTICQLVVGQ
                VLEPLRLKHPSLCMLHYMDDLLLAASSHDGLEAAGEEVISTLERAGFTISPDKVQREP
                GVQYLGYKLGSTYVAPVGLVAEPRIATLWDVQKLVGSLQWLRPALGIPPRLMGPFYEQ
```

LRGSDPNEAREWNLDMKMAWREIVRLSTTAALERWDPALPLEGAVARCEQGAIGVLGQ
GLSTHPRPCLWLFSTQPTKAFTAWLEVLTLLITKLRASAVRTFGKEVDILLLPACFRE
DLPLPEGILLALKGFAGKIRSSDTPSIFDIARPLHVSLKVRVTDHPVPGPTVFTDASS
STHKGVVVWREGPRWEIKEIADLGASVQQLEARAVAMALLLWPTTPTNVVTDSAFVAK
MLLKMGQEGVPSTAAAFILEDALSQRSAMAAVLHVRSHSEVPGFFTEGNDVADSQATF
QAYPLREAKDLHTALHIGPRALSKACNISMQQAREVVQTCPHCNSAPALEAGVNPRGL
GPLQIWQTDFTLEPRMAPRSWLAVTVDTASSAIVVTQHGRVTSVAVQHHWATAIAVLG
RPKAIKTDNGSCFTSKSTREWLARWGIAHTTGIPGNSQGQAMVERANRLLKDRIRVLA
EGDGFMKRIPTSKQGELLAKAMYALNHFERGENTKTPIQKHWRPTVLTEGPPVKIRIE
TGEWEKGWNVLVWGRGYAAVKNRDTDKVIWVPSRKVKPDITQKDEVTKKDEASPLFAG
ISDWIPWEDEQEGLQGETASNKQERPGEDTLAANES"

mat_peptide 2736..5420
/product="reverse transcriptase beta subunit"

mat_peptide 2736..2786
/product="reverse transcriptase alpha subunit"

CDS 613..2718
/note="gag-Pr76 polyprotein precursor"
/codon_start=1
/protein_id="AAB59932.1"
/db_xref="GI:210175"
/translation="MEAVIKVISSACKTYCGKTSPSKKEIGAMLSLLQKEGLLMSPSD
LYSPGSWDPITAALSQRAMILGKSGELKTWGLVLGALKAAREEQVTSEQAKFWLGLGG
GRVSPPGPECIEKPATERRIDKGEEVGETTVQRDAKMAPEETATPKTVGTSCYHCGTA
IGCNCATASAPPPPYVGSGLYPSLAGVGEQQGQGGDTPPGAEQSRAEPGHAGQAPGPA
LTDWARVREELASTGPPVVAMPVVIKTEGPAWTPLEPKLITRLADTVRTKGLRSPITM
AEVEALMSSPLLPHDVTNLMRVILGPAPYALWMDAWGVQLQTVIAAATRDPRHPANGQ
GRGERTNLNRLKGLADGMVGNPQGQAALLRPGELVAITASALQAFREVARLAEPAGPW
ADIMQGPSESFVDFANRLIKAVEGSDLPPSARAPVIIDCFRQKSQPDIQQLIRTAPST
LTTPGEIIKYVLDRQKTAPLTDQGIAAAMSSAIQPLIMAVVNRERDGQTGSGGRARGL
CYTCGSPGHYQAQCPKKRKSGNSRERCQLCNGMGHNAKQCRKRDGNQGQRPGKGLSSG
PWPGPEPPAVSLAMTMEHKDRPLVRVILTNTGSHPVKQRSVYITALLDSGADITIISE
EDWPTDWPVMEAANPQIHGIGGGIPMRKSRDMIELGVINRDGSLERPLLLFPAVAMVR
GSILGRDCLQGLGLRLTNL"

mat_peptide 613..1137
/product="P19 protein"

mat_peptide 1144..1329
/product="P10 protein"

mat_peptide 1330..2049
/product="P27 protein"

mat_peptide 2077..2343
/product="P12 protein"

mat_peptide 2344..2715
/product="P15 protein"

mat_peptide 2344..2346
/product="P15 protein"

exon <613..2343
/note="gag-pol-Pr180 polyprotein precursor"
/number=1

exon 613..2343
/note="gag-pol-Pr180 polyprotein precursor; putative"
/number=1

CDS join(613..630,914..1270)
/note="transcriptional activator protein"
/codon_start=1
/protein_id="AAB59931.1"
/db_xref="GI:210172"
/translation="MEAVIKGEGGSLPQVRSASRNQQRSGESTKGRKWEKQLCSEMRR
WRRRKRPHLKPLAHPAIIAEQLLAVIAPQPRLLLLLMWGVVCILPWRGWESSRARGVT
HLRGRNSQGRSQGMRVRLLGRP"

exon <613..630
/note="env-Pr95 polyprotein precursor"

```
                     /number=1
     intron          631..7286
                     /note="21S (src) intron A"
     intron          631..5310
                     /note="env-Pr95 intron A"
     intron          631..913
                     /note="trn-act intron A"
     exon            914..>1270
                     /note="transcriptional activator protein,  [5]"
                     /number=2
     exon            2736..5423
                     /note="gag-pol-Pr180 polyprotein precursor; putative"
                     /number=2
     exon            2736..>5423
                     /note="gag-pol-Pr180 polyprotein precursor"
                     /number=2
     exon            5311..>7098
                     /note="env-Pr95 polyprotein precursor"
                     /number=2
     CDS             7362..8942
                     /note="src-p60 phosphoprotein"
                     /codon_start=1
                     /protein_id="AAB59935.1"
                     /db_xref="GI:210176"
                     /translation="MGSSKSKPKDPSQRRHSLEPPDSTHHGGFPASQTPDETAAPDAH
                     RNPSRSFGTVATEPKLFWGFNTSDTVTSPQRAGALAGGVTTFVALYDYESWTETDLSF
                     KKGERLQIVNNTEGDWWLAHSLTTGQTGYIPSNYVAPSDSIQAEEWYFGKITRRESER
                     LLLNPENPRGTFLVRKSETAKGAYCLSVSDFDNAKGPNVKHYKIYKLYSGGFYITSRT
                     QFGSLQQLVAYYSKHADGLCHRLANVCPTSKPQTQGLAKDAWEIPRESLRLEAKLGQG
                     CFGEVWMGTWNDTTRVAIKTLKPGTMSPEAFLQEAQVMKKLRHEKLVQLYAVVSEEPI
                     YIVIEYMSKGSLLDFLKGEMGKYLRLPQLVDMAAQIASGMAYVERMNYVHRDLRAANI
                     LVGENLVCKVADFGLARLIEDNEYTARQGAKFPIKWTAPEAALYGRFTIKSDVWSFGI
                     LLTELTTKGRVPYPGMVNREVLDQVERGYRMPCPPECPESLHDLMCQCWRKDPEERPT
                     FKYLQAQLLPACVLEVAE"
     LTR             9291..9625
                     /note="3' LTR"
     repeat_region   9525..9545
                     /note="3' terminal repeat"
BASE COUNT     2300 a   2421 c   2761 g   2143 t
ORIGIN
        1 atgtagtctt atgcaatact cctgtagtct tgcaacatgc ttatgtaacg atgagttagc
       61 aatatgcctt acaaggaaag aaaaggcacc gtgcatgccg attggtggta gtaaggtggt
      121 acgatcgtgc cttattagga aggtatcaga cgggtctaac atggattgga cgaaccactg
      181 aattccgcat cgcagagata ttgtatttaa gtgcctagct cgatacaata aacgccattt
      241 taccattcac cacattggtg tgcacctggg ttgatggccg gaccgtcgat tccctaacga
      301 ttgcgaacac ctgaatgaag cagaaggctt catttggtga ccccgacgtg atagttaggg
      361 aatagtggtc ggccacagac ggcgtggcga tcctgccctc atccgtctcg cttattcggg
      421 gagcggacga tgaccctagt agaggggct gcggcttagg agggcagaag ctgagtggcg
      481 tcggagggag ctctactgca gggagcccag ataccctacc gagaactcag agagtcgttg
      541 gaagacggga aggaagcccg acgactgagc agtccacccc aggcgtgatt ctggtcgccc
      601 ggtggatcaa gcatggaagc cgtcataaag gtgatttcgt ccgcgtgtaa aacctattgc
      661 gggaaaacct ctccttctaa gaaggaaata ggggccatgt tgtccctctt acaaaaggaa
      721 gggttgctta tgtctccctc agacttatat tccccggggt cctgggatcc cattaccgcg
      781 gcgctatccc agcgggctat gatacttggg aaatcgggag agttaaaaac ctggggattg
      841 gttttggggg cattgaaggc ggctcgagag gaacaggtta catctgagca agcaaagttt
      901 tggttgggat taggggggagg gagggtctct cccccaggtc cggagtgcat cgagaaacca
      961 gcaacggagc ggcgaatcga caaaggggag gaagtgggag aaacaactgt gcagcgagat
     1021 gcgaagatgg cgccggagga aacggccaca cctaaaaccg ttggcacatc ctgctatcat
     1081 tgcggaacag ctattggctg taattgcgcc acagcctcgg ctcctcctcc tccttatgtg
     1141 gggagtggtt tgtatccttc cctggcgggg gtgggagagc agcagggcca ggggggtgac
     1201 acacctccgg gggcggaaca gtcaagggcg gagccagggc atgcgggtca ggctcctggg
```

```
1261 ccggccctga ctgactgggc aagggtcagg gaggagcttg cgagtactgg tccgcccgtg
1321 gtggccatgc ctgtagtgat taagacagag ggacccgctt ggacccctct ggagccaaaa
1381 ttgatcacaa gactggctga tacggtcagg accaagggct tacgatcccc gattactatg
1441 gcagaagtgg aagcgcttat gtcctccccg ctgctgccgc atgacgtcac gaatctaatg
1501 agagttattt tagggcctgc cccatatgcc ttatggatgg acgcttgggg agtccaactc
1561 cagacagtta tagcggcagc cactcgcgac ccccgacacc cagcgaacgg tcaagggcgg
1621 ggggaacgga ctaatttgaa tcgcttaaag ggcttagctg atgggatggt gggcaaccca
1681 cagggtcagg ccgcattatt aagaccgggg gaattggttg ctattacggc gtcggctctc
1741 caggcgttta gagaggttgc ccggctggcg gaacctgcag gtccatgggc ggacatcatg
1801 cagggaccat ctgagtcctt tgttgatttt gccaatcggc ttataaaggc ggttgagggg
1861 tcagatctcc cgccttccgc gcgggctccg gtgatcattg actgctttag gcagaagtca
1921 cagccagata ttcagcagct tatacggaca gcaccctcca cgctgaccac cccaggagag
1981 ataattaaat atgtgctaga caggcagaag actgcccctc ttacggatca aggcatagcc
2041 gcggccatgt cgtctgctat ccagccctta attatggcag tagtcaatag agagagggat
2101 ggacaaactg ggtcgggtgg tcgtgcccga gggctctgct acacttgtgg atccccggga
2161 cattatcagg cgcagtgccc gaaaaaacgg aagtcaggaa acagccgtga gcgatgtcag
2221 ttgtgtaacg ggatgggaca caacgctaaa cagtgtagga agcgggatgg caaccagggc
2281 caacgcccag gaaaaggtct ctcttcgggg ccgtggcccg gccctgagcc acctgccgtc
2341 tcgttagcga tgacaatgga acataaagat cgccccttgg ttagggtcat tctgactaac
2401 actgggagtc atccggtcaa acagcgttcg gtgtatatca ccgcgctgtt ggactctgga
2461 gcggacatca ctattatttc agaggaggat tggcccaccg attggccagt gatggaggcc
2521 gcgaacccgc agatccatgg gataggaggg ggaattccca tgcgaaaatc tcgtgacatg
2581 atagagttgg gggttattaa ccgagacggg tctttggagc gacccctgct cctcttcccc
2641 gcagtagcta tggttagagg gagtatccta ggaagagatt gtctgcaggg cctagggctc
2701 cgcttgacaa atttataggg agggccactg ttctcactgt tgcgctacat ctggctattc
2761 cgctcaaatg gaagccagac cacacgcctg tgtggattga ccagtggccc ctccctgaag
2821 gtaaacttgt agcgctaacg caattagtgg aaaaagaatt acagttagga catatagaac
2881 cttcacttag ttgttggaac acacctgtct tcgtgatccg gaaggcttcc gggtcttacc
2941 gcttactgca tgatttgcgc gctgttaacg ccaagcttgt tccttttggg gccgtccaac
3001 agggggcgcc agttctctcc gcgctcccgc gtggctggcc cctgatggtc ttagacctca
3061 aggattgctt cttttctatc cctcttgcgg aacaagatcg cgaagctttt gcatttacgc
3121 tcccctctgt gaataaccag gccccgctc gaagattcca atggaaggtc ttgccccaag
3181 ggatgacctg ttctcccact atctgtcagt tggtagtggg tcaggtactt gagcccttgc
3241 gactcaagca cccatctctg tgcatgttgc attatatgga tgatcttttg ctagccgcct
3301 caagtcacga tgggttggaa gcggcagggg aggaggttat cagtacattg gaaagagccg
3361 ggttcactat ttcgcctgat aaggtccaga gggagcccgg agtacaatat cttgggtaca
3421 agttaggcag tacgtatgta gcacccgtag gcctggtagc agaacccagg atagccacct
3481 tgtgggatgt tcaaaagctg gtggggtcac ttcagtggct tcgcccagcg ttaggaatcc
3541 cgccacgact gatgggcccc ttctatgagc agttacgagg gtcagatcct aacgaggcga
3601 gggaatggaa tctagacatg aaaatggcct ggagagagat cgtacggctt agcaccactg
3661 ctgccttgga acgatgggac cctgccctgc ctctggaagg agcggtcgct agatgtgaac
3721 agggggcaat aggggttttg ggacagggac tgtccacaca cccaaggcca tgcttgtggt
3781 tattctccac ccaacccacc aaggcgttta ctgcttggtt agaagtgctc acccttttga
3841 ttactaagct acgtgcttcg gcagtgcgaa cctttggcaa ggaggtcgat atcctcctgt
3901 tgcctgcatg ctttcgggag gaccttccgc tcccagaggg gatcctgtta gcccttaagg
3961 ggtttgcagg aaaaatcagg agtagtgaca cgccatctat ttttgacatt gcgcgtccac
4021 tgcatgtttc tctgaaagtg agggttaccg accaccctgt gccgggaccc actgtcttta
4081 ctgacgcctc ctcaagcacc cataaggggg tggtagtctg gagggagggc ccaaggtggg
4141 agataaaaga aatagctgat ttgggggcaa gtgtacaaca actggaagca cgcgctgtgg
4201 ccatggcact tctgctgtgg ccgacaacgc ccactaatgt agtgactgac tccgcgtttg
4261 ttgcgaaaat gttactcaag atgggacagg agggagtccc gtctacagcg gcggctttta
4321 ttttagagga tgcgttaagc caaaggtcag ccatggccgc cgttctccac gtgcggagtc
4381 attctgaagt gccagggttt ttcacagaag gaaatgacgt ggcagatagc caagccacct
4441 tccaagcgta tcccttgaga gaggctaaag atcttcatac cgctctccat attggacccc
4501 gcgcgctatc caaagcgtgt aatatatcta tgcagcaggc tagggaggtt gttcagacct
4561 gcccgcattg taattcagcc cctgcgttgg aggccggagt aaaccctagg ggtttgggac
4621 ccctacagat atggcagaca gactttacgc ttgagcctag aatggccccc cgttcctggc
4681 tcgctgttac tgtggatacc gcctcatcag cgatagtcgt aactcagcat ggccgtgtca
4741 catcggttgc tgtacaacat cattgggcca cggctatcgc cgttttggga agaccaaagg
4801 ccataaaaac agataatggg tcctgcttca cgtctaaatc cacgcgagag tggctcgcga
4861 gatgggggat agcacacacc accgggattc cgggtaattc ccagggtcaa gctatggtag
```

```
4921 agcgggccaa ccggctcctg aaagatagga tccgtgtgct tgcggagggg gacggcttta
4981 tgaaaagaat ccccaccagc aaacaggggg aactattagc caaggcaatg tatgccctca
5041 atcactttga gcgtggtgaa aacacgaaaa caccgataca aaaacactgg agacctaccg
5101 ttcttacaga aggacccccg gttaaaatac gaatagagac aggggagtgg gaaaaaggat
5161 ggaacgtgct ggtctgggga cgaggttatg ccgctgtgaa aaacagggac actgataagg
5221 ttatttgggt accctctcga aaagttaaac cggacatcac ccaaaaggat gaggtgacta
5281 agaaagatga ggcgagccct ctttttgcag gcatttctga ctggataccc tgggaagacg
5341 agcaagaagg actccaagga gaaaccgcta gcaacaagca agaaagaccc ggagaagaca
5401 cccttgctgc caacgagagt taattatatt ctcattattg gtgtcctggt cttgtgtgag
5461 gttacggggg taagagctga tgttcactta ctcgagcagc cagggaacct ttggattaca
5521 tgggccaacc gtacaggcca aacggatttc tgcctctcta cacagtcagc cacctcccct
5581 tttcaaacat gtttgatagg tatcccgtct cctatttccg aaggtgattt taagggatat
5641 gtttctgata caaattgctc cactgtggga actgaccggt tagtcttgtc agccagcatt
5701 accggcggcc ctgacaacag caccaccctc acttatcgaa aggtttcatg cctgctgtta
5761 aagctgaacg tctccatgtg ggatgagcca cctgaactgc agctgctagg ttcccagtct
5821 ctccctaacg ttactaacat tactcaggtc tctggcgtgg ccgggggatg tgtatatttc
5881 gccccaaggg ccactggcct gtttttaggt tggtctaaac aaggtctctc gcggttcctc
5941 ctccgtcacc cctttacctc cacctctaac tccacggaac cgttcacggt ggtgacagcg
6001 gatagacaca atctttttat ggggagtgag tactgtggtg catatggcta cagattttgg
6061 gaaatatata actgctcaca gactaggaat acttaccgct gtggagacgt gggaggtact
6121 ggcctccctg aaacctggtg cagaggaaaa ggaggtatat gggttaatca atcaaaggaa
6181 attaatgaga cagagccgtt cagttttact gcgaactgta ctggcagtaa tttgggtaat
6241 gtcagcggat gttgcggaga accaatcacg attctcccac taggggcatg gatcgacagt
6301 acgcaaggta gtttcactaa accaaaagcg ctaccacccg caattttcct catttgtggg
6361 gatcgcgcat ggcaaggaat tcccagtcgt ccggtagggg gcccctgcta tttaggcaag
6421 cttaccatgt tagcacccaa ccatacagat attctcaaaa tacttgctaa ttcgtcgcgg
6481 acaggtataa gacgtaaacg aagcgtctca cacctggatg atacatgctc agatgaagta
6541 cagctttggg gtcctacagc aagaatcttt gcatctatct tagccccggg ggtagcagct
6601 gcgcaagcct taagagaaat tgagagacta gcctgttggt ccgttaaaca ggctaacttg
6661 acaacatcac tcctcgggga cttattggat gatgtcacga gtattcgaca cgcggtcctg
6721 cagaaccgag cggctattga cttcttgctt ctagctcacg gccatggctg tgaggacgtt
6781 gccggaatgt gttgtttcaa tctgagtgat cacagtgaat ctatacagaa gaagttccag
6841 ctaatgaaga aacatgtcaa taagatcggc gtggacagcg acccaatcgg aagttggctg
6901 cgagggatat tcgggggaat aggggaatgg gccgttcatc tgctaaaagg actgcttttg
6961 gggcttgtag ttattttatt gctactggtg tgcctgcctt gcctttaca atttgtgtct
7021 agtagtattc gaaagatgat taatagttca atcaactatc atactgaata caggaagatg
7081 cagggcggag cagtctagag ctcagttata ataatcctgc gaatcgggct gtaacggggc
7141 aaggcttgac cgaggggact ataacatgta taggcgaaaa gcggggtctc ggttgtaacg
7201 cgcttaggaa gtcccctcga ggtatggcag atatgctctt gcatagggga aaaaaatgta
7261 gtcttaatat tgtctgtgtg ctgcaggagc taagctgact ctgctggtgg cctcgcgtac
7321 cactgtggcc aggcggtagc tgggacgtgc agccgaccac catggggagc agcaagagca
7381 agcctaagga ccccagccag cgccggcaca gcctggagcc acccgacagc acccaccacg
7441 ggggattccc agcctcgcag acccccgacg agacagcagc ccccgacgca caccgcaacc
7501 ccagccgctc cttcgggacc gtggccaccg agcccaagct cttctggggc ttcaacactt
7561 ctgacaccgt cacgtcgccg cagcgtgccg gggcactggc tggcggcgtc accactttcg
7621 tggctctcta cgactacgag tcctggactg aaacggactt gtccttcaag aaaggagaac
7681 gcctgcagat tgtcaacaac acggaaggtg actggtggct ggctcattcc ctcactacag
7741 gacagacggg ctacatcccc agtaactatg tcgcgccctc agactccatc caggctgaag
7801 agtggtactt tgggaagatc actcgtcggg agtccgagcg gctgctgctt aaccccgaaa
7861 acccccgggg aaccttcttg gtccggaaga gcgagacggc aaagggtgcc tattgcctct
7921 ccgtttctga ctttgacaac gccaaggggc ccaatgtgaa gcactacaag atctacaagc
7981 tgtacagcgg cggcttctac atcacctcac gcacacagtt cggcagccta cagcagctgg
8041 tggcctacta ctccaaacat gctgatggct tgtgccaccg cctggccaac gtctgcccca
8101 cgtccaagcc ccagacccag ggactcgcca aggacgcgtg ggaaatcccc cgggagtcgc
8161 tacggctgga ggcgaagctg gggcagggct gctttggaga ggtctggatg gggacctgga
8221 acgacaccac cagagtggcc ataaagactc tgaagcccgg caccatgtcc ccggaggcct
8281 tcctgcagga agcccaagtg atgaagaagc tccggcatga gaagctggtt cagctgtacg
8341 cagtggtgtc ggaagagccc atctacatcg tcattgagta catgagcaag gggagcctcc
8401 tggatttcct gaagggagag atgggcaagt acctgcggct gccacagctc gtcgatatgg
8461 ctgctcagat tgcatccggc atggcctatg tggagagaat gaactacgtg caccgagacc
8521 tgcgggcggc caacatcctg gtgggggaga acctggtgtg caaggtggct gacttcgggc
```

```
     8581 tggcacgcct catcgaggac aacgagtaca cagcacggca aggtgccaag ttccccatca
     8641 agtggacagc ccccgaggca gccctctatg gccggttcac catcaagtcg gatgtctggt
     8701 ccttcggcat cctgctgact gagctgacca ccaagggccg ggtgccatac ccagggatgg
     8761 tcaacaggga ggtgctggac caggtggaga ggggctaccg catgccctgc ccgcccgagt
     8821 gccccgagtc gctgcatgac ctcatgtgcc agtgctggcg gaaggaccct gaggagcggc
     8881 ccacctttaa gtacctgcag gcccagctgc tccctgcttg tgtgttggag gtcgctgagt
     8941 aagtacgagg cgtgacctac aattgctcaa ataatgcttc tgtagaaatt gtttagcatt
     9001 aggcgtcctg cgttgctccg cgatgtacgg gtcaggtata atgtgcagtt tgactgaggg
     9061 gaccatgatg tgtataggcg tcaagcgggg cttcggttgt acgcggatag gaatcccctc
     9121 aggacaattc tgcttggaat atgatggcgt cttccctgtt ttgcccttag actattcgag
     9181 ttgcctctgt ggattagggc tggaggcagc acggatagtc tgatggccaa ataaggcagg
     9241 caagacagct atttgtaact gcgaaatacg cttttgcata gggaggggga aatgtagtct
     9301 tatgcaatac tcctgtagtc ttgcaacatg cttatgtaac gatgagttag caatatgcct
     9361 tacaaggaaa gaaaaggcac cgtgcatgcc gattggtggt agtaaggtgg tacgatcgtg
     9421 ccttattagg aaggtatcag acgggtctaa catggattgg acgaaccact gaattccgca
     9481 tcgcagagat attgtattta agtgcctagc tcgatacaat aaacgccatt ttaccattca
     9541 ccacattggt gtgcacctgg gttgatggcc ggaccgtcga ttccctaacg attgcgaaca
     9601 cctgaatgaa gcagaaggct tcatt
//
```

THE JOURNAL OF BIOLOGICAL CHEMISTRY
1993 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 268, No. 4, Issue of February 5, pp. 2674-2683, 1993
Printed in U.S.A.

Enzymatic Analysis of Two HIV-1 Reverse Transcriptase Mutants with Mutations in Carboxyl-terminal Amino Acid Residues Conserved among Retroviral Ribonucleases H*

(Received for publication, June 15, 1992)

Silke Volkmann‡, Birgitta Maria Wöhrl‡§, Margaret Tisdale¶, and Karin Moelling‡‖

From the ‡Max-Planck-Institut für Molekulare Genetik, Abteilung Schuster, Ihnestrasse 73, D-1000 Berlin 33, Germany and the ¶Wellcome Research Laboratories, Molecular Sciences Department, South Eden Park Road, Beckenham, Kent BR3 3BS, United Kingdom

The reverse transcriptase (RT) of HIV-1 has been mutagenized within the carboxyl-terminal domain which harbors the RNase H. Two amino acids highly conserved among all 14 known RT sequences but not in th bacterial RNase H have been mutagenized resulting in the mutant proteins N494D and Q475E. Th y were expressed as recombinant proteins, purifi d, and analyzed for their *in vitro* properties in comparison to the p66 homodimeric wild-type and a previ usly described H539N mutant. The N494D mutant cl sely resembles the wild-type RNase H, exhibits an end nuclease activity and a processive RNase H activity, gives rise to small RNA hydrolysis products, and a ts in concert with the RT. The Q475E mutant is more d fective and resembles the H539N mutant, exhibits a r tard d endonuclease activity and an impaired 3′ → 5′ processive RNA cleavage activity, gives rise to predominantly larger RNA hydrolysis products, is less processive in the presence of competitor substrate, and is d fective in its ability to hydrolyze the polypurine tract and homopolymeric hybrids. Short homopolym ric stretches cause a pausing of the RT of wild-type and mutants which results in a coordinated action of th RNase H. Pausing of the RT correlates with RNase H cleavages about 20 nucleotides behind the point of synthesis. The defects of the mutant enzymes can be interpreted on the basis of the known crystallography data.

Retroviruses replicate by means of the viral-coded reverse transcriptase (RT)[1] which converts the single-stranded RNA genome into double-stranded DNA and allows integration into the cellular genome. The process of reverse transcription leads to production of the long-terminal repeats , the viral promotor-enhancer element (for review, see Varmus and Swanstrom (1984), Varmus (1987), and Goff (1990)). The RT is a multi-functional enzyme that possesses RNA- and DNA-dependent DNA polymerase activities and RNase H (Baltimore, 1970; Temin and Mitzutami, 1970; Moelling *et al.*, 1971; Gilboa *et al.*, 1979). The latter activity specifically cleaves RNA in RNA-DNA hybrids (for review, see Crouch (1990), Wintersberger (1990), and Moelling (1992)). For initiation of cDNA synthesis, the RT requires RNA primers. It uses a tRNA primer for minus-strand DNA synthesis and the polypurine tract (PPT) for plus-strand DNA synthesis. The RNase H is required for several stages during viral replication. It removes the RNA template so that a DNA copy can be converted to double-stranded DNA (Watson, 1979). Furthermore, the RNase H removes the primers, the tRNA, and the PPT RNA (Omer and Faras, 1982; Furfine and Reardon, 1991; Champoux *et al.*, 1984; Huber and Richardson, 1990). The function of the RNase H has been identified as involving endo- as well as exonuclease activity (Krug and Berger, 1989; Oyama *et al.*, 1989; Dudding *et al.*, 1990; Schatz *et al.*, 1990; Wöhrl and Moelling, 1990). The exonuclease activity does not give rise predominantly to mononucleotides and has therefore been defined by DeStefano *et al.* (1991b) as a partially processive 3′ → 5′ endonuclease.

Catalyzing all these functions requires their efficient coordination. Some efforts have been made to localize the enzyme functions and to study their interdependence. From these studies the RT present in HIV-1 virions was shown to be a heterodimer p66/p51 with a common NH_2 terminus (Di-Marzo-Veronese *et al.*, 1986; Lowe *et al.*, 1988). By means of recombinant techniques it can be expressed as p66 homodimer that harbors both polymerase and RNase H activity (Hansen *et al.*, 1987, 1988) or as carboxyl-terminally truncated p51 protein that lacks RNase H activity and has a strongly reduced polymerase function (Lori *et al.*, 1988; Hansen *et al.*, 1988; Tisdale *et al.*, 1989). The RNase H domain of HIV-1 resides in the carboxyl-terminal part of the RT protein (Hansen *et al.*, 1988).

The carboxyl-terminal region has been recovered as a p15 protein from a RT enzyme preparation from virus particles and shown to harbor RNase H activity (Hansen *et al.*, 1988). Renaturation of p15 RNase H in activated gel assays allowed detection of RNase H activity, whereas the recombinant p15 was inactive (Becerra *et al.*, 1990; Hostomsky *et al.*, 1991). A combination of recombinant p15 with p51 leads to reconstitution of RNase H activity (Hostomsky *et al.*, 1991). Recently, p15 RNase H has been isolated as a processing product from the recombinant p66 RT and has been reported to exhibit polymerase activity, which is an unexpected result (Hafkemeyer *et al.*, 1991). Recent data indicate that the RT also exhibits double-stranded RNA cleaving activity designated as RNase D (Ben-Artzi *et al.*, 1992).

Although the RNase H and the polymerase functions can

* This work was supported by Grant FKZII-019-86 from the Bundesminister für Forschung und Technologie (BMFT), Dr. Mildred Scheel Stiftung F66, the Deutsche Forschungsgemeinschaft, and Max Planck Gesellschaft. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

This work is dedicated to Dr. Heinz Schuster on the occasion of his 65th birthday.

§ Present address: Case Western Reserve University, School of Medicine, Cleveland, OH 44106.

‖ To whom correspondence should be addressed. Tel.: 30-8307254; Fax: 30-8307383.

[1] The abbreviations used are: RT, reverse transcriptase including RNase H; HIV-1, human immunodeficiency virus 1; PPT, polypurine tract; DTT, dithiothreitol; TBE, Tris-borate-EDTA.

be separated with some retrovirus RTs (Tanese and Goff, 1988; Prasad and Goff, 1989; Hansen *et al.*, 1988; Lei and Verma, 1978), an interdependence for both enzyme activities has been assumed. Different models describing either cooperative or uncoupled mechanisms for the RT and the RNase H have been proposed (Oyama *et al.*, 1989; Schatz *et al.*, 1990; Wöhrl and Moelling, 1990; DeStefano *et al.*, 1991a).

Recently the crystal structure of the HIV-1 p66/p51 heterodimer complexed with an inhibitor has been published (Kohlstaedt *et al.*, 1992). The crystal structures of the RNase H from *Escherichia coli* as well as the HIV-1 RNase H have been solved successfully (Katayanagi *et al.*, 1990, 1992; Yang *et al.*, 1990; Davies *et al.*, 1991). Both enzymes share several highly conserved amino acids and exhibit highly related structures.

Several of these conserved residues have been mutated in studies on functional analysis of the HIV-1 RT. The results indicated that D443 and D498 of the HIV-1 RT belong to the RNase H active center (Mizrahi *et al.*, 1990). H539 apparently is essential for binding to the RNA-DNA substrate (Wöhrl *et al.*, 1991). Mutations of H539 and E478 have been shown to abolish virus infectivity (Schatz *et al.*, 1989; Tisdale *et al.*, 1991).

Two mutant enzymes are characterized in the present study which have been mutagenized in amino acids highly conserved among all known retroviral RNase H domains but not in the *E. coli* RNase H. These mutants are being characterized with respect to their enzymatic functions in comparison to the corresponding wild-type and the previously analyzed mutant H539N. The defects of these mutants are discussed with respect to their location and the crystal structure.

EXPERIMENTAL PROCEDURES

Materials—Restriction enzymes were purchased from BRL, Berlin, Germany. *Hae*III-restricted pBR322 DNA, phage T4 polynucleotide kinase, *E. coli* RNA polymerase, calf intestine phosphatase, RNase-free DNase I, M13 mp9 single-stranded DNA, 2′,3′-dideoxynucleoside triphosphates (ddNTPs), 2′-deoxynucleoside triphosphates (dNTPs), and poly(rA)·$(dT)_{10}$ (base ratio 1:1) were from Boehringer, Mannheim, Germany. RNasin was purchased from Promega Biotech, Heidelberg, Germany. Poly(rA), poly(dC), and the vector pTZ19R were from Pharmacia LKB, Freiburg, Germany. Radiolabeled compounds [γ-^{32}P]ATP (3,000 Ci/mmol), [α-^{32}P]UTP (>400 Ci/mmol), and [α-^{32}P]GTP (>400 Ci/mmol) were from Amersham Buchler, Braunschweig, Germany. The RNA transcription kit was purchased from Stratagene, Heidelberg, Germany. DNA sequencing kit was from United States Biochemical Corp., Cleveland, OH. Deoxyoligonucleotides were synthesized on an Applied Biosystems oligonucleotide synthesizer and were purified by high pressure liquid chromatography prior to use.

Protein Purification—Expression and purification of HIV-1 RT p66 wild-type and mutant enzymes were performed as described previously (Hansen *et al.*, 1987; Wöhrl *et al.*, 1991). One unit of purified HIV-1 RT (60 ng), wild-type, or N494D mutant protein, catalyzes the incorporation of 13 pmol of [^{3}H]TMP into acid-insoluble products in 10 min at 37 °C with poly(rA)·$(dT)_{10}$ as template/primer. The specific activity of the Q475E mutant protein is about 10% reduced. The H539N mutant protein reaches about 70% of the wild-type activity. One unit RNase H activity degrades 10 fmol of [α^{32}P] RNA/M13DNA hybrid in 1 min at 37 °C as determined by acid-soluble radioactivity using filter assays. 60 ng of wild-type or N494D mutant protein correspond to one unit. The specific activity for the Q475E and H539N mutant enzymes is reduced by 20%. The RNase H activity depends on the properties of the substrate.

DNA Manipulations—Restriction endonuclease cleavage, DNA isolation, end-labeling of DNA, and transformations were performed as described by Maniatis *et al.* (1989). *In vitro* transcription reactions were performed as described by the manufacturer (Stratagene).

Plasmid Construction—Plasmids pKJ11 and pKJ2 are derivatives of plasmid pTZ P8 described previously (Wöhrl and Moelling, 1990). The insert of pTZ P8 was partially digested by *Eco*RV resulting two *Hin*dIII-*Eco*RV, fragments which were cloned into the vector plasmid pTZ19R via the *Hin*dIII and *Sma*I sites yielding pKJ2 and pKJ11.

In Vitro Transcription—Plasmids pTZ P8, pKJ2, and pKJ11 were linearized with *Eco*RI and transcribed with T7 RNA polymerase as described previously for pTZ P8 (Wöhrl and Moelling, 1990). Thus pKJ2 yielded an RNA of 134 bases, pKJ11 yielded an RNA of 213 bases (see Fig. 5*c*), and pTZ P8 yielded an RNA of 534 bases in length (see Fig. 1). All RNAs include the PPT.

Preparation of 5′-End-labeled RNA—*In vitro* transcribed RNA of pKJ2 or pKJ11 (~20 pmol) was dephosphorylated with calf intestine phosphatase at 50 °C for 1 h. The reaction was terminated by the addition of stop buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 200 mM NaCl, 0.5% SDS). After phenol extraction and ethanol precipitation, the RNA was end-labeled with 50 μCi of [γ-^{32}P]ATP and 30 units of T4 polynucleotide kinase for 1 h at 37 °C. The sample was again extracted with phenol, ethanol-precipitated, resuspended in 20 μl of urea loading buffer (7 M urea in 1 × TBE, 0.1% each xylene cyanol and bromphenol blue), and electrophoresed on a 6% polyacrylamide-TBE-urea-gel. After autoradiography the band containing the RNA was excised from the gel and eluted in 400 μl of elution buffer (0.5 M CH_3COONH_4, 1 mM EDTA, 0.1% SDS) at 37 °C overnight, phenol-extracted, precipitated in 2.5 volumes of ethanol, and washed twice with 80% ethanol. The dried precipitate was dissolved in water and an aliquot measured by scintillation counting. The specific activity was approximately 1–2 × 10^4 cpm/pmol.

RNA/DNA Hybridization—*In vitro* transcribed RNA was hybridized with a 2–3-fold molar excess of a specific deoxyoligonucleotide in a 50 mM Tris-HCl, pH 8.0, buffer containing 80 mM KCl, 1 mM DTT, and 1 mM EDTA to produce the hybrids pKJ2/Y (see Figs. 1 and 2), pKJ2/X (see Figs. 1 and 3), pKJ11/N (see Figs. 1, 4, and 5 (*a–c*)), and pTZ P8/S (see Figs. 1 and 6). For analysis of cDNA elongation, [γ-^{32}P]ATP end-labeled deoxyoligonucleotide was used, and for analysis of the RNase H cleavage products, [γ-^{32}P]ATP 5′-end-labeled RNA. The mixture was incubated for 3 min at 90 °C, then cooled slowly to room temperature.

Preparation of [α-^{32}P]RNA/DNA Hybrids—In a total volume of 250 μl, 15 μg of poly(dC) was incubated for 1 h at 37 °C with 30 units of *E. coli* RNA polymerase, 40 nM GTP, and 100 μCi of [α-^{32}P]GTP in 50 mM Tris-HCl buffer containing 100 mM KCl, 5 mM $MgCl_2$, 5 mM DTT, and 1 mM $MnCl_2$. For preparing the [α-^{32}P]RNA/M13 DNA hybrid, 5 μg of M13 mp9 single-stranded DNA was incubated in a total volume of 250 μl for 1 h with 10 units of RNA polymerase, 100 nM ATP, CTP, and GTP, respectively, and 30 nM UTP and 50 μCi of [α-^{32}P]UTP in 40 mM Tris-HCl, pH 8.0, buffer containing 8 mM $MgCl_2$ and 100 mM KCl. After phenol extraction and ethanol precipitation, the precipitate was dried, dissolved in 200 μl of TE buffer, and purified from unincorporated nucleotides by gel filtration using a Sephadex G50 column. 500-μl fractions were collected, and the fractions containing the hybrid were detected by scintillation counting.

RNase H Cleavage Assays—RNase H cleavage of hybrids consisting of *in vitro* transcribed RNA and a DNA oligonucleotide and RNase H reactions coupled with limited cDNA elongation were performed as described previously (Wöhrl and Moelling, 1990; Wöhrl *et al.*, 1991), using 5′-end-labeled RNA. Unless otherwise stated about 10 pmol (10,000 cpm) of radioactively labeled RNA were used for each reaction. Aliquots were analyzed by electrophoresis on appropriate polyacrylamide-TBE-urea gels.

Degradation Products—For determination of the degradation products of the RNase H reaction, 60 ng of wild-type or mutant enzymes were incubated in 50 μl of RNase H buffer (50 mM Tris-HCl, pH 8.0, 40 mM NaCl, 1 mM $MgCl_2$, 2 mM DTT) with 500,000 cpm of either [α-^{32}P]poly(rG)·poly(dC) (~20 ng) or [α-^{32}P]RNA/M13-DNA hybrid at 37 °C. 10-μl aliquots of the wild-type and N494D mutant enzyme were taken and after 20 min, mixed with 5 μl of urea loading buffer, heated for 2 min at 95 °C, and put on ice. Aliquots from the Q475E and H539N mutant enzymes were taken after 20 and 90 min. Aliquots of the [α-^{32}P]poly(rG)·poly(dC) degradation products were analyzed on 20% polyacrylamide-TBE-urea gels, and aliquots of the [α-^{32}P] RNA/M13-DNA reaction products were analyzed on 16% polyacrylamide-TBE-urea gels. The amount of degradation were determined by scanning microdensitometry of the autoradiograms.

RNase H Cleavage during RNA-dependent cDNA Synthesis—For analyzing the RNase H cleavage products during RNA-dependent cDNA synthesis, 30 ng of substrate pKJ11/N (see Fig. 1) containing 5′-end-labeled RNA was incubated with 24 ng of wild-type or mutant enzymes and 20 units of RNasin in 23.5 μl of RT-standard-buffer. After 5 min at 37 °C RNA-dependent cDNA synthesis was started by addition of $MgCl_2$ (8 mM) and dNTPs (100 μM) to give a final reaction volume of 25 μl. Reactions were stopped after 1, 5, 10, 30, and 60 min with 5 μl of 0.5 M EDTA. Finally all samples were ethanol-precipitated

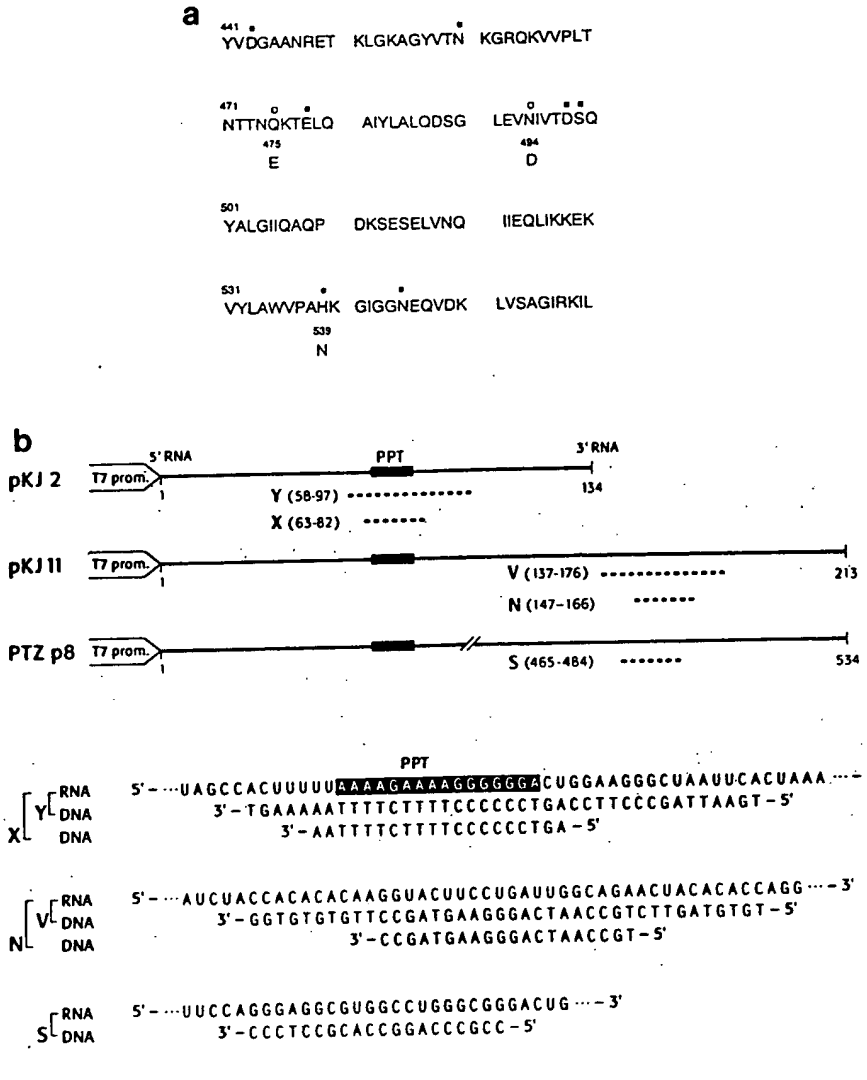


FIG. 1. **Sequences of RNase H and hybrid substrates.** *Panel a*, amino acid sequence (441–560) of the carboxyl-terminal RNase H of HIV-1. Seven amino acids, indicated by *black squares*, are conserved among retroviral and bacterial enzymes. Two additional ones, indicated by *open squares* are conserved among all known retroviral sequences. Positions of point mutations are Q475, N494, and H539. Mutants are Q475E, N494D, and H539N. *Panel b*, the RNA/DNA substrates used in this study are shown. The polypurine tract (PPT) is indicated as a *black bar*. The RNA was transcribed *in vitro* from a T7 promoter, using the vector pTZ p8 and derivatives. The RNAs are designated as pTZ p8, pKJ11, and pKJ2 according to the *in vitro* transcription templates. The DNA oligonucleotides named *X*, *Y*, *V*, *N*, and *S* were synthesized *in vitro* and are indicated as *dotted lines* (*top*) or listed as nucleotides (*bottom*).

and analyzed on a 10% TBE-urea-polyacrylamide gel. In a second experiment carried out and analyzed as described above the cDNA elongation was analyzed by using the same substrate pKJ11/N containing a 5′-end-labeled deoxyoligonucleotide.

Measurement of Polymerization during DNA Synthesis in the Presence of Competitor Substrate—cDNA elongation was analyzed under conditions of either absence or excess of competitor RNA. 24 ng of wild-type or mutant enzymes were mixed with 40 ng of hybrid pKJ11/N and 20 units of RNasin in 21.5 μl of RT-standard-buffer. After a 5-min preincubation period at 37 °C poly(rA)·$(dT)_{10}$ (5 μg, base ratio 1:1, as indicated by the supplier) was added. DNA elongation was started by addition of $MgCl_2$ (8 mM) and dNTPs (100 μM), either together with poly(rA)·$(dT)_{10}$ or 30, 60, 120 and 180 s after addition of the competitor poly(rA)·$(dT)_{10}$. Reactions were stopped after 10 min with 5 μl of 0.5 M EDTA. The assay conditions were tested in a control experiment carried out without $MgCl_2$ and dNTPs, or with adding the competitor RNA before the labeled substrate pKJ11/N and without adding competitor RNA with all four enzymes (data not shown). Finally all probes were precipitated with ethanol and analyzed on a 10% polyacrylamide-TBE-urea gel.

Measurement of RNase H Cleavage during DNA Synthesis in the Presence of Competitor Substrate—For analyzing the RNase H cleavage products created under assay conditions described for the cDNA synthesis with competition, 24 ng of wild-type or mutant enzymes were mixed for preincubation with 10 ng of pKJ11/N hybrid containing 5′-end-labeled RNA. Immediately or 30, 60, and 180 s after addition of competitor substrate, the reaction was started and carried out as described.

RNA-dependent cDNA Synthesis—RNA-dependent cDNA synthesis was carried out with template/primer pTZ P8/S (see Fig. 1) containing the specific end-labeled deoxyoligonucleotide. In a volume of 17.8 μl of 50 mM Tris-HCl, pH 8.0, buffer containing 40 mM KCl, 20 mM NaCl, 2 mM DTT, and 1 mM EDTA, 3 nM template/primer was incubated with 24, 12, 6, or 1.5 ng of wild-type or mutant enzymes for 5 min at 37 °C. The reaction was started by addition of 250 μM dNTPs and 6 mM $MgCl_2$ to give a final volume of 20 μl. After 10 min at 37 °C, the probes were incubated at 95 °C for 2 min, precipitated with ethanol, and analyzed on an 8% polyacrylamide-TBE-urea gel. For analyzing elongation stops, the sequence of the template was determined in parallel using ^{32}P-end-labeled deoxyoligonucleotide S as sequencing primer and pTZ P8 plasmid DNA as sequencing template. The sequencing reaction was carried out according to the manufacturer's instructions.

RESULTS

RNase H Cleavage Reactions—Various functional parameters of two homodimeric p66 mutant enzymes, Q475E and N494D were compared with the homodimeric wild-type protein and the recently described H539N mutant enzyme (Wöhrl *et al.*, 1991). The RNase H activity was first analyzed in the absence of dNTPs. The reactions were performed with 5′-end-labeled *in vitro* transcribed pKJ2 RNA hybridized to a synthetic 40-mer deoxyoligonucleotide Y resulting in a RNA/DNA hybrid that spanned the region of the polypurine tract (PPT) (Fig. 1). The PPT is the *in vivo* start point for the plus-strand DNA synthesis of retroviruses (Champoux *et al.*, 1984; Resnick *et al.*, 1984). The wild-type and mutant enzymes were purified as described (Wöhrl *et al.*, 1991) and tested with various NaCl concentrations for their RNase H

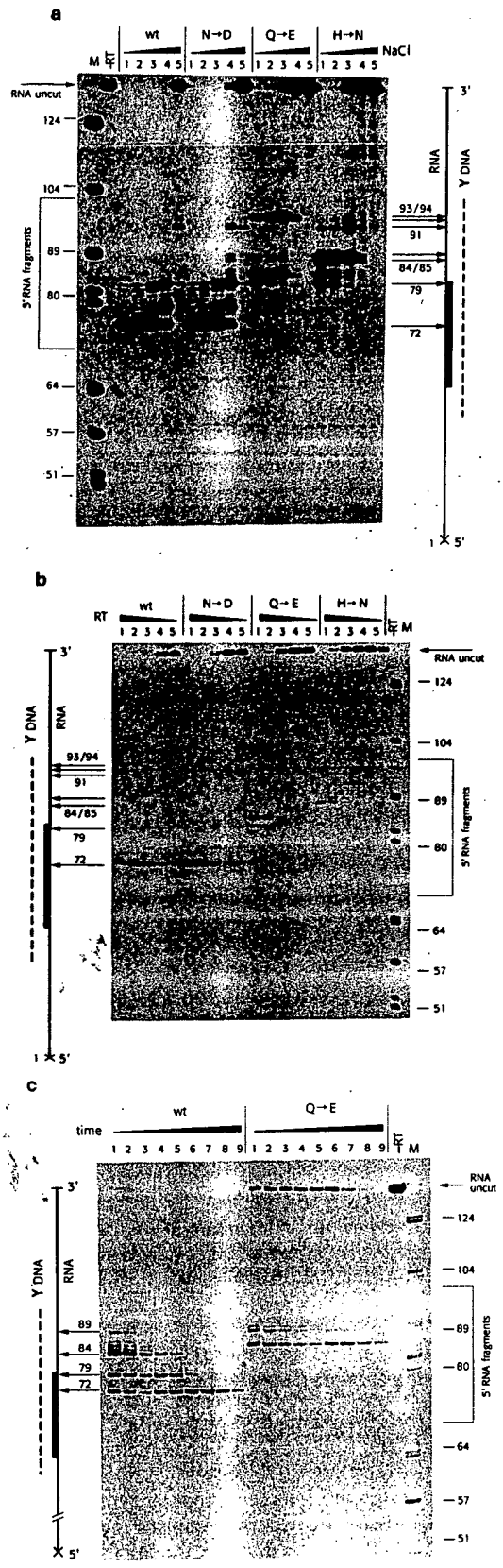


FIG. 2. **Parameters affecting the cleavage patterns of hybrid pKJ2/Y with the p66 wild-type and mutant enzymes.** *a*, dependence on salt concentrations. Hybrid pKJ2/Y end-labeled with ^{32}P (indicated by *cross*) were incubated with 6 ng of wild-type or mutant RT enzymes as indicated for 30 min at 37 °C in the standard reaction mixture at final NaCl concentrations of 25, 65, 105, 145, and 225 mM NaCl (*lanes 1–5*) and analyzed on 10% polyacrylamide-TBE-urea gels. The *arrows* point to the cleavage sites. *−RT* indicates incubation of hybrid without enzyme; *M* represents *Hae*III fragments of pBR322 as molecular weight markers. *Numbers* refer to nucleotides. *b*, dependence on enzyme concentration. Hybrid pKJ2/Y was treated with 60, 25, 3.75, 0.94, and 0.45 ng of wild-type or mutant enzymes as indicated (*lanes 1–5*) for 30 min at 37 °C and then analyzed on a 10% polyacrylamide-TBE-urea gel. For other explanations see above. *c*, time course analysis of the cleavage reaction with hybrid pKJ2/Y. The substrate was incubated with wild-type or Q475E mutant enzyme in RNase H buffer with a final NaCl concentration of 65 mM. The standard reaction mixture was scaled up to 100 μl. After 15 and 30 s and 1, 2, 5, 10, 20, 30, and 60 min, 10-μl portions were taken from the reaction mixture heated for 3 min to 80 °C and analyzed on a 10% polyacrylamide-TBE-urea gel. For other explanations see above.

TABLE I

Characteristics of degradation products analyzed on a homopolymeric hybrid A and a heteropolymeric hybrid B

500,000 cpm (~20 ng) of [α-^{32}P]poly(rG)·poly(dC) (hybrid A) was treated with 60 ng of wild-type or mutant enzymes as described under "Experimental Procedures." The same procedure was carried out with an [α-^{32}P]RNA/M13-DNA hybrid (hybrid B). The numbers refer to the length (nucleotides) of the predominant hydrolysis products.

Enzyme	Length of main degradation products		Percentage degraded	
	Hybrid A	Hybrid B	Hybrid A	Hybrid B
			%	
Wild-type	7–4	12–6	>90	>90
N494D	7–4	12–6	>90	>90
Q475E	10–8	18–15	≈10	≈80
H539N	10–8	22–18	≈20	≈80

activities (Fig. 2*a*). The RNA cleavage products were analyzed on 10% polyacrylamide-TBE-urea gels and visualized by autoradiography. The results obtained with the PPT containing hybrid pKJ2/Y (Fig. 2*a*), indicate that the cleavage pattern observed with the wild-type arises by superposition of two effects, an endonucleolytic cleavage of the RNA and a further shortening by a 3′ → 5′ processive RNase H activity. The wild-type and one of the mutant enzymes, N494D, generate a prominent reaction product of 72 nucleotides in length. In the case of the two other mutant enzymes the initial endonucleolytic cleavage site at nucleotide 94/93 is preserved. The subsequent processive 3′ → 5′ RNA cleavage activity is reduced leading to major intermediate cleavage sides (nucleotide 85/84 and 79) at the 3′ side of the PPT. The PPT is most highly conserved. Increase of salt concentration, shown in Fig. 2*a*, leads to strong inhibition of the RNase H activities of both mutants, Q475E and H539N, whereas the RNase H of the wild-type and N494D mutant is only inhibited at the highest salt concentration. The experiment shown in Fig. 2*a* has also been performed using a hybrid without PPT, pKJ11/V (compare Fig. 1). Under these conditions the subsequent processive RNA cleavage activity only of the Q475E and H539N mutants is inhibited (data not shown).

A further test was performed with decreasing RT concentrations (Fig. 2*b*). Again, the wild-type and the N494D mutant enzyme produce nearly identical cleavage patterns, whereas those obtained with mutant Q475E and H539N, are longer and the main cleavage products of the two mutant proteins differ from each other. As has already been shown for the H539N mutant (Wöhrl *et al.*, 1991), the endonuclease activity of the Q475E mutant appears to be impaired in the PPT region, which results in preferential cleavage sites 3′ of the PPT. To obtain a better resolution of the RNase H activity, a time course experiment was performed with the Q475E mutant in comparison with the wild-type. The incubation periods ranged from 15 s to 60 min (Fig. 2*c*). The wild-type exhibits endonuclease and subsequent processive 3′ → 5′

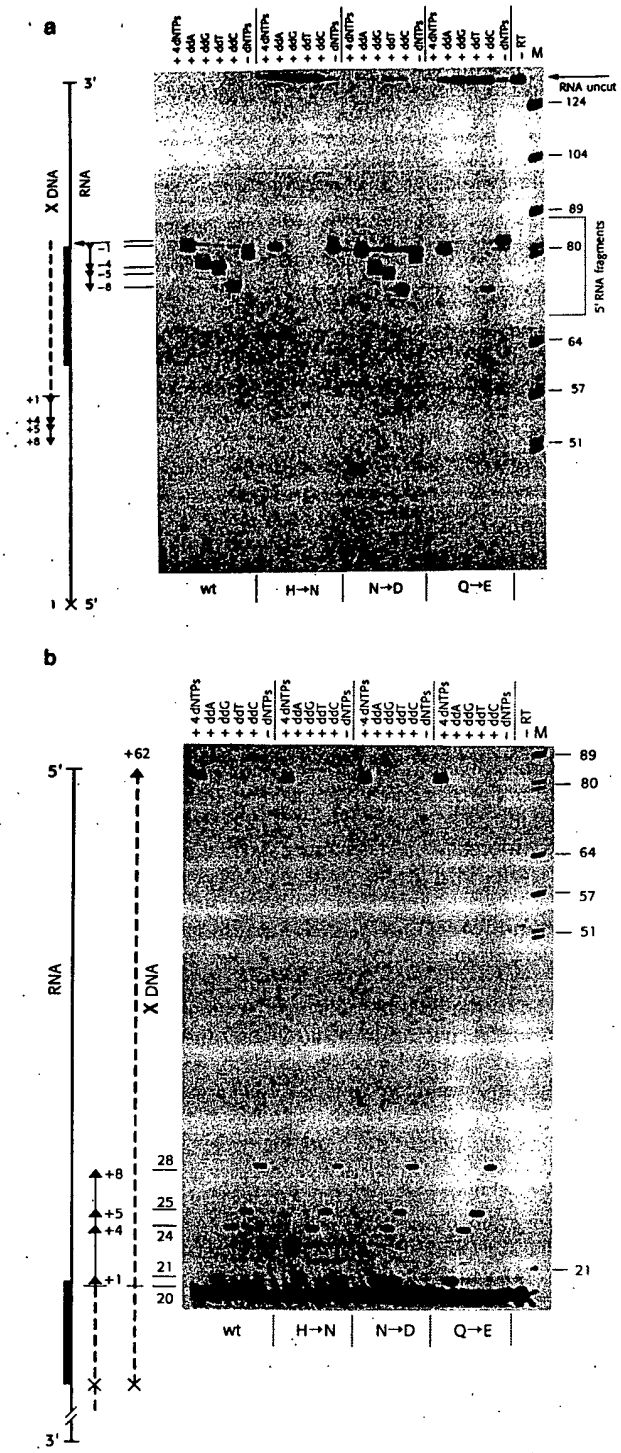


FIG. 3. **RNase H cleavage combined with limited cDNA synthesis.** *a*, RNA cleavage pattern obtained in a limited cDNA elongation assay was carried out with pKJ2/X hybrid containing 5′-end-labeled RNA. In a total volume of 10 μl (containing 50 mM Tris-HCl, pH 8.0, 80 mM NaCl, 1 mM $MgCl_2$, and 2 mM DTT), 12 ng of wild-type or mutant enzyme were incubated with 20 ng of labeled RNA/DNA hybrid. The assay was performed in the absence or presence of four dNTPs or three dNTPs plus one ddNTP as described (Wöhrl and Moelling, 1990; Wöhrl *et al.*, 1991). Incubation was for 30 min at 37 °C. The 3′ and 5′ RNA cleavage products were separated on a 10% polyacrylamide TBE-urea gel. Addition of ddA, ddG, ddC, and ddT as indicated allow elongation of the DNA primer by 1, 4, 5, and 8 deoxynucleotides, respectively. *+4 dNTPs* and *−dNTPs* indicate presence of all four or absence of deoxynucleotides, respectively. *−RT* indicates a control in the absence of RT, *M* represents *Hae*III fragments of pBR322 as molecular weight markers, and *numbers* refer to nucleotides. Numbers of nucleotides hydrolyzed and synthesized are indicated by − and + signs. *b*, evidence for cDNA elongation as control for *a*. The reactions were carried out as described in *a*, except that the RNA was unlabeled and the DNA oligonucleotide in the hybrid ^{32}P-end-labeled. 10% polyacrylamide-TBE-urea gel was used. For explanation of numbers, see *a*.

RNA cleavage activity in a time-dependent fashion, whereas the mutant is impaired in its initial endonuclease as well as the subsequent 3′ → 5′ RNase H activity. Again the PPT region is most highly conserved. This behavior is similar to that of the H539N mutant enzyme described previously (Wöhrl *et al.*, 1991) and can be similarly interpreted as an inhibition of the latter 3′ → 5′ RNase H activity. Our results indicate that the first endonucleolytic cut and the subsequent processive RNase H activity can function independently of each other. They are also dependent on the sequence of the hybrid because two of the three mutant enzymes (H539N, Q475E) are especially inhibited in cleaving the RNA in the PPT region.

Degradation Products—To determine the degradation products of the RNase H reaction, homo- and heteropolymeric RNA/DNA hybrids were used. The homopolymer consisted of poly(dC) and a radioactively labeled poly(rG) strand synthesized with *E. coli* RNA polymerase *in vitro*. The heteropolymer was derived from single-stranded M13 DNA. Synthesis of the complementary RNA was also performed by *E. coli* RNA polymerase using [^{32}P]UTP as label. Both RNA/DNA hybrids were incubated for 20 min. The results are shown in Table I; 90% of both hybrids were hydrolyzed by the wild-type and N494D mutant within 20 min, and the distribution of the degradation products was similar in size and amount. The H539N and Q475E mutant enzymes only degraded 10–20% of the homopolymer and 80% of the heteropolymer hybrid. Whereas the wild-type and N494D mutant enzyme degraded both hybrids up to 4-mers, the Q475E and H539N mutant enzymes created larger degradation products, 8-mers with the homopolymeric and 14-mers with the heteropolymer hybrid. Longer incubation periods (90 min) did not lead to smaller RNA fragments (data not shown). With all enzymes, mutant and wild-type, the predominant degradation products are always larger than mononucleotides. These findings are in agreement with the results of Mizrahi (1989) and DeStefano *et al.* (1991b); however, Starnes and Cheng (1989) primarily observed mono-, di-, and trimers with a poly(rG)·poly(dC) substrate.

RNase H Cleavage Combined with Limited cDNA Synthesis—After determination of the mode of action of the RNase H activity in the absence of dNTPs, a second approach was taken under conditions allowing limited cDNA elongation. The limitation was achieved by replacing one dNTP by a dideoxynucleotide, which leads to chain termination (Sanger *et al.*, 1977). The assay was carried out as described (Wöhrl and Moelling, 1990), except that 5′-end-labeled RNA was used. The short hybrid pKJ2/X was chosen to enable substrate binding only for 1 enzyme molecule/hybrid molecule. Analysis of the 5′ RNA fragments gives information about the processing of the RNA in a RNA/DNA hybrid in the presence of restricted primer extension (Fig. 3*a*). The wild-type as well as the N494D mutant enzyme lead to endonucleolytic cleavage at the 3′ side of the PPT. Primer extensions by 1, 4, 5, and 8 nucleotides (Fig. 3*b*) correspond to decrease in size of the RNA fragments by 1, 4, 5, and 8 ribonucleotides (Fig. 3*a*), suggesting that an initial endonucleolytic cleavage event was followed by a 3′ → 5′ processive RNase H activity. The wild-type and the N494D mutant enzyme are able to process the substrate nearly to completion (compare the amount of uncut RNA, Fig. 3*a*). The Q475E and H539N mutant enzymes are strongly inhibited in their RNase H

activity and incapable of cleaving the PPT in the RNA/DNA hybrid. For the H539N mutant enzyme this feature has already been described (Wöhrl *et al.*, 1991). Control reactions in the presence or absence of all four dNTPs result in endonucleolytic cuts at the 3′ side out of the PPT. The RNA-dependent DNA polymerase activity of all four enzymes exhibit no impaired use of the intact PPT region as primer for cDNA elongation (Fig. 3*b*).

RNase H Cleavage during RNA-dependent cDNA Synthesis—How does the RNase H activity process RNA in the presence of cDNA synthesis? To answer this question a time course experiment with incubation periods ranging from 1 to 60 min was carried out with the hybrid pKJ11/N in the presence of dNTPs to allow primer extension. The RNA degradation fragments were analyzed using 5′-end-labeled RNA. The time course was carried out in parallel with 5′-end-labeled DNA primer to determine the efficiency of the RNA-dependent DNA polymerization. The wild-type enzyme shows the strongest RNase H activity and processes the RNA nearly completely giving rise to predominant degradation fragments of 89 and 14–18 ribonucleotides in length. The N494D mutant enzyme creates only long 5′ RNA fragments during short incubation periods (1 and 5 min, see Fig. 4*a*, *lanes 1* and *2*), which are further processed with increasing time ultimately leading to results similar to those for the wild-type. This mutant enzyme behaves like a retarded wild-type RNase H activity. The Q475E mutant enzyme behaves differently, leading to a sequential appearance of intermediate-sized 5′ RNA degradation fragments, and only a few very short ones (smaller than 18 ribonucleotides) after long incubation periods. The H539N mutant enzyme is even more impaired in its RNase H activity than the Q475E mutant, leaving large amounts of the RNA uncut. The Q475E and H539N mutant enzymes are also less efficient in RNA-dependent DNA polymerase activity and reach full-length DNA transcripts only after longer incubation periods (Fig. 4*b*).

Measurement of RNase H Cleavage during DNA Synthesis in the Presence of Competitor Substrate—The results presented in Fig. 4 show prominent DNA stops resulting 65- and 89-mer DNA fragments (Fig. 4*b*) and a prominent 89-mer RNA fragment for the wild-type and the N494D mutant enzyme (Fig. 4*a*). In order to understand the occurrence of these reaction products and for a better resolution of the interaction between RNase H and RNA-dependent DNA polymerase activity, a further assay was performed under conditions that limit the reaction to one RT molecule using excess of challenger substrate during primer extension. Under these conditions a RT molecule is allowed to function until it detaches from the substrate. For analysis of the RNA hydrolysis of the pKJ11/N hybrid containing 5′-end-labeled RNA, the hybrid was preincubated with the wild-type or one of the mutant enzymes to allow an enzyme-substrate complex to form. Poly(rA)·$(dT)_{10}$ (5 μg) was added to each sample as challenger substrate, and DNA synthesis was started by addition of $MgCl_2$ and dNTPs after increasing time periods (Fig. 5, *lanes 1–4*). Under these conditions the largest amount of RNA remains uncut. The wild-type leads to two dominant RNA fragments of about 100 ribonucleotides. The N494D mutant enzyme leads to similar fragments. These disappear in both cases in the absence of competitor (*lane A*) resulting in 89-mer RNA hydrolysis products presumably by the exonucleolytically action of the RNase H. Control reactions were performed to analyze the nature of these predominant RNase H reaction products (89-mers). The resulting bands in *A* are resistant to subsequent addition of exogenous *E. coli* RNase H (*lanes B*), indicating that they do not consist of RNA/DNA hybrids. A possible explanation for the mode of action of the wild-type and the N494D mutant enzyme at this

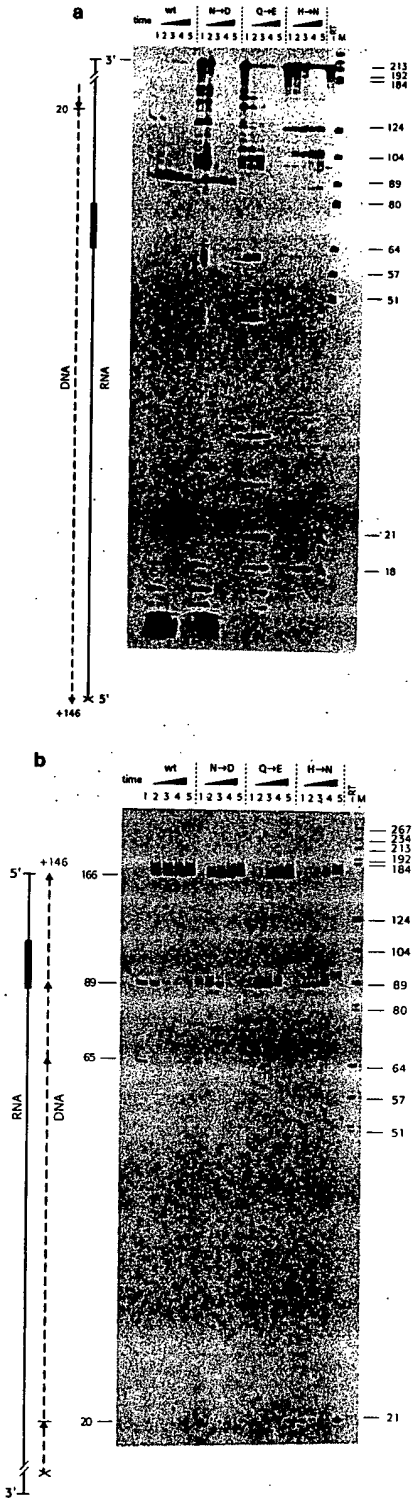


FIG. 4. **Time course of RNase H cleavage and RNA-dependent cDNA synthesis.** *a*, RNase H cleavage pattern of hybrid pKJ11/N containing a 5′-end-labeled RNA moiety is shown in a time course experiment (for details see "Experimental Procedures"). Wild-type or mutant enzymes (24 ng each) were incubated for 1 min (*lanes 1*), 5 min (*lanes 2*), 15 min (*lanes 3*), 30 min (*lanes 4*), or 60 min (*lanes 5*). Reaction products were analyzed on a 10% polyacrylamide-TBE-urea gel. *M* indicates deoxyoligonucleotide markers. *Lane −RT* represents a control without RT. *b*, the same experiment as described in *a* was carried out, except that the hybrid pKJ11/N contained a 5′-end-labeled DNA oligonucleotide (indicated by *cross* at *dotted arrow*) instead of end-labeled RNA.

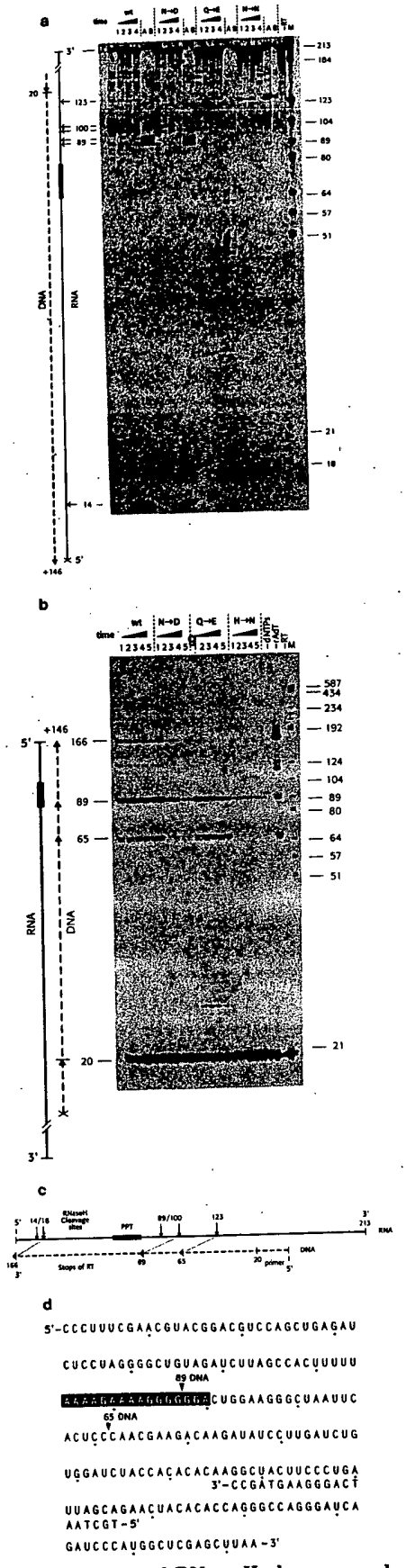


FIG. 5. **Measurement of RNase H cleavage and polymerization during DNA synthesis in the presence of competitor substrate.** *a*, 10 ng of the hybrid pKJ11/N (see Fig. 1) containing 5′-end-labeled RNA was incubated with 24 ng of wild-type or mutant enzymes. After 5 min at 37 °C poly(rA)·(dT)$_{10}$ (5 μg) was added as competitor substrate. Synthesis was started by the addition of 6 mM $MgCl_2$ and 100 μM dNTPs at various time points, 0 s (*lanes 1*), or after 30 s (*lanes 2*), 60 s (*lanes 3*), or 180 s (*lanes 4*). For comparison the assays were carried out without competitor (*lanes A*) and were additionally treated with *E. coli* RNase H (*lanes B*). *−RT* represents a control without RT. *M* indicates deoxyoligonucleotide markers (*numbers*). *Numbers* and *arrows* indicate RNase H cleavage sites. Samples were analyzed on a 10% polyacrylamide-TBE urea gel. *b*, hybrid pKJ11/N (40 ng) containing end-labeled DNA oligonucleotide N was incubated with 24 ng of wild-type or mutant enzymes. After 5 min at 37 °C poly(rA)·(dT)$_{10}$ (5 μg) as competitor substrate was added. The reaction was started by the addition of 6 mM $MgCl_2$ and 100 μM dNTPs after 0 s (*lanes 1*), after 30 s (*lanes 2*), 60 s (*lanes 3*), 120 s (*lanes 4*), and 180 s (*lanes 5*). The lanes indicated with *−dNTPs* and *−RT* represent negative controls carried out without dNTPs or without RT. For comparison the reaction was also carried out using wild-type enzyme without competitor substrate (indicated as *−rAdT*). *M* indicates marker. Probes were analyzed on a 10% polyacrylamide-TBE-urea gel. *c* represents a scheme of the RNase H cleavage sites (*perpendicular arrows*) and stops of DNA elongation of the hybrid pKJ11/N. *Dotted lines* connect RNase H cleavage sites and stops of DNA elongation. *d*, the sequences of the RNA template and the DNA primer N are listed. The PPT is delineated in *black*. *Arrowheads* and *numbers* point to the DNA elongation stops.

site may be the proximity of the PPT, which may block the polymerase activity due its particular structure (see below). The Q475E and H539N mutant enzymes create a different RNA degradation pattern in the presence of competitor. They exhibit highly impaired RNase H activities leading to longer RNA cleavage products. In the absence of competitor (*A*), those mutant enzymes perform sporadic cleavages of the full-length hybrid but create no 14-mers as the wild-type or the N494D mutant.

As can be concluded from the scheme in Fig. 5*c*, the large RNA cleavage products (larger than 147-mers) result from hydrolysis of the initial hybrid pKJ11/N. The shorter RNA fragments detected mainly with the wild-type must thereby arise as a consequence of RNA-dependent DNA synthesis, which is limited to only one round. This reaction clearly demonstrates a coupling of RNase H and RNA-dependent DNA polymerase activity.

Measurement of Polymerization during DNA Synthesis in the Presence of Competitor Substrate—In order to elucidate the mode of coupling of RNase H and polymerase activity it was essential to determine the DNA synthesis reactions. Although the use of 5′-end-labeled RNA allows visualization of the effects of the RNase H (Fig. 5*a*), the cDNA polymerase activity can be determined using 5′-end-labeled DNA primer (Fig. 5*b*). Full-length DNA transcripts of 166 nucleotides in length are only obtained by the wild-type and the N494D mutant enzyme. Although the N494D mutant enzyme is easily competed for during preincubation, as can be concluded from the decreasing amounts of reaction products with increasing time periods (Fig. 5, *a* and *b*), the wild-type protein shows no such effect. All four enzymes create cDNA elongation stops resulting in DNA transcripts of 65 and a second dominant one of 89 nucleotides in length, determined by sequencing reactions (data not shown). These elongation stops lead to RNase H cleavage sites of 123 and 100 ribonucleotides, which are 22 to 23 nucleotides apart from the DNA synthesis stop. This is illustrated in Fig. 5*c*. Only the wild-type protein allows full-length cDNA (166-mer) and corresponding 14–18-mer RNA fragments. For the Q475E mutant enzyme a number of very short DNA transcripts can be detected, suggesting an impairment of the processivity of its RT. The RNA fragments are correspondingly large. To analyze the nature of the cDNA elongation stops, the primary template RNA sequences were analyzed. The two prominent elongation stops coincide with

short stretches of homopolymeric RNA sequences (Fig. 5*c*), one of them with the PPT.

RNA-dependent cDNA Elongation—In order to verify the significance of homopolymeric stretches for elongation stops of the RT, primer-extension was analyzed without competition using another hybrid, pTZ P8/S as template/primer. The primer-extension was carried out in this case with decreasing amounts of wild-type and mutant enzyme. A sequencing reaction with the template and end-labeled DNA oligonucleotide S as primer was carried out in parallel. The results shown in Fig. 6 confirm the importance of short homopolymeric, G- or C-rich stretches on the RNA for pausing of the RT in agreement with the results shown in Fig. 5 and a previous analysis (Dudding *et al.*, 1991). The stops (numbered 1 through 6) are most clearly detected with the Q475E mutant.

DISCUSSION

Seven amino acids are conserved among retroviral and bacterial RNase H; D443, E478, D498, S499, H539, N545, and D549. Two additional ones, N494 and E475, are invariant among all 14 known retroviral RT sequences (Doolittle *et al.*, 1989). We have performed site-directed mutagenesis of these two amino acids N494D and Q475E of the homodimeric p66 HIV-1 RT. The properties of their RNase H functions were determined in comparison to the wild-type protein and a recently described H539N mutant enzyme (Wöhrl *et al.*, 1991). The functional properties of the mutant enzymes were determined and are discussed with regard to x-ray crystallography data of the heterodimeric p66/p51 HIV-1 RT (Kohlstaedt *et al.*, 1992), the retroviral (Davies *et al.*, 1991) and the bacterial RNase H (Katayanagi *et al.*, 1990, 1992; Yang *et al.*, 1990).

The wild-type RNase H activity exhibits endonuclease and a separable processive RNase H activity which cleaves the RNA in the 3′ → 5′ direction. The N494D mutant enzyme behaves similarly to the wild-type, whereas Q475E and H539N mutant enzymes only perform endonucleolytic cuts using the hybrid pKJ11/V. From this experiment it can be concluded that the endonuclease activity and the processive RNase H activity can be separated from each other. The Q475 and H539N mutant enzymes are apparently impaired in this latter function. For the H539N mutant we have shown recently that the 3′ → 5′ processive RNase H cleavage activity is not lost completely but severely reduced (Wöhrl *et al.*, 1991). This is also the case for the Q475E mutant protein, as evidenced by the time course analysis (Fig. 2*c*).

The wild-type and mutant enzymes analyzed here consist only of the p66 polypeptide and lack p51 processing products. It has been described recently that the p66/p51 heterodimeric RT is somewhat more processive than its homodimeric counterpart (Huang *et al.*, 1992). In a previous study the H539N mutant enzyme has been analyzed in comparison to the homo- and heterodimeric wild-type enzyme whereby only minor differences in the mode of action of the two wild-types were detectable (Wöhrl *et al.*, 1991). Therefore we do not expect major differences between homo- and heterodimeric mutant enzymes, even though this remains to be shown.

From our data the PPT appears to be a special hindrance for the polymerase as well as the RNase H activity. The wild-type and the N494D mutant enzyme create larger remaining hybrids (18–20 nucleotides, Fig. 3*a*) if the PPT sequence itself serves as substrate for the RNase H. The Q475E and H539N mutant enzymes do not cleave within the PPT. During reverse transcription the RNA-dependent polymerase activity leads to a dominant elongation stop within the PPT (89-mer fragments, Figs. 4*a* and 5*a*). In the case that the polymerase pauses within the PPT the RNase H of the wild-type and the N494D mutant enzymes cleave the random RNA sequences

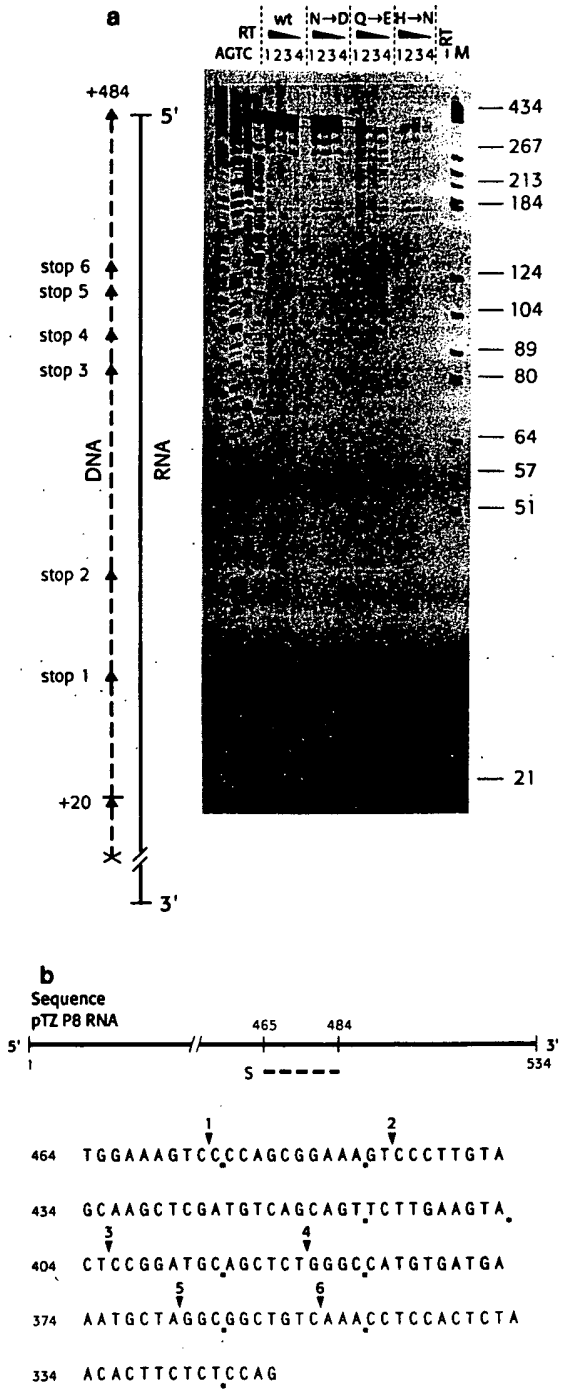


FIG. 6. **Pattern of DNA transcripts created during cDNA synthesis on an RNA template.** *a*, decreasing concentrations of wild-type or mutant enzymes, 24 ng (*lanes 1*), 12 ng (*lanes 2*), 6 ng (*lanes 3*), and 1.5 ng (*lanes 4*) were incubated at 37 °C with hybrid (3 nM) consisting of the RNA template pTZ P8 and the ^{32}P-end-labeled DNA oligonucleotide S under standard DNA synthesis conditions. The full-length DNA transcript (484-mer) and the dominant elongation stops of the RNA-dependent DNA polymerase reaction are indicated as *stops 1–6* on the *left*. The *four left lanes* (*A*, *G*, *T*, and *C*) represent a sequence reaction to determine the sequence of the transcripts, which is listed below. *−RT* represents a negative control without enzyme. *M* indicates DNA markers. Samples were analyzed on an 8% polyacrylamide-TBE-urea gel. *b*, schematic drawing of the RNA template pTZ P8, primer S, and the sequence of the first part of the DNA transcript. Stops are indicated by *numbered arrowheads* corresponding to the stops in *a*.

3′ of the PPT leading to a remaining hybrid-region 12–14 nucleotides in length (Fig. 5*a*, *lanes A* and *B*).

It is not only the special PPT that influences the functional properties of the wild-type and mutant enzymes; analysis of the RNA degradation products of a homopolymeric hybrid leads to similar results. Wild-type and N494D mutant enzymes can hydrolyze more than 90% of the homopolymeric hybrid, whereas the two other mutant enzymes leave 80–90% uncut (Table I). A heteropolymeric hybrid is hydrolyzed with similar efficiency by all four enzymes (Table I). The size of the hydrolysis products created by the wild-type and mutant enzymes was always larger than mononucleotides, which leads De Stefano and colleagues (1991b) to define the RNase H as a processive 3′ → 5′ endonuclease.

It has been controversial to what extent the RNase H activity is coupled to the RT (De Stefano *et al.*, 1991a; Wöhrl *et al.*, 1991; Oyama *et al.*, 1989). We have addressed this question by two approaches: by using limited DNA synthesis due to presence of dideoxynucleotides (Fig. 3) and, in a second approach, by applying competitor as described by De Stefano *et al.* (1991) (Fig. 5). In the presence of chain terminators, a coordination of RT and RNase H is observed for the wild-type and the N494D mutant enzyme. Elongation of the cDNA strand corresponds to the hydrolysis of the RNA strand in a constant distance of 18–20 base pairs (Fig. 3). A coordinated interplay of polymerase and RNase H is also detected in the presence of competitor, which limits the action of a single RT molecule to one hybrid until it detaches. Distinct cDNA elongation stops occur under these conditions (Fig. 5*b*), which lead to prominent RNA degradation fragments (Fig. 5*a*). The DNA stops and RNase H cleavage sites are about 20 nucleotides apart. A similarly coordinated coupling between polymerase and RNase H activity occurred not only with a PPT-containing hybrid but also with a hybrid lacking the PPT (data not shown). This distance is somewhat longer than that deduced from some published biochemical studies (Oyama *et al.*, 1989; Furfine and Reardon, 1991) but is in good agreement with our previous result and crystal structure data published recently (Kohlstaedt *et al.*, 1992). A duplex RNA/DNA hybrid of about 20 base pairs fits between the two active sites in the structure of heterodimeric HIV-RT with model built A-form RNA/DNA hybrid.

The coordinated action between the RNase H and the polymerase coincides with a pausing of polymerase. This occurs either due to the presence of one dideoxynucleotide (Fig. 3), or at short homopolymeric stretches of G and C (Figs. 5 and 6). A similar behavior can be concluded from the data reported by Dudding *et al.* (1991). DNA elongation stops are most prominent with the Q475E mutant enzyme.

A close look at the RNA-dependent DNA synthesis shows that only the wild-type and the N494D mutant enzyme can reach full-length cDNAs. These correlate with generation of short RNA degradation fragments (Fig. 5, *a* and *b*). The H539N mutant is most highly impaired in the overall DNA synthesis (Figs. 5*b* and 6*a*). The Q475E mutant exhibits a reduced processivity in comparison to the wild-type and the N494D mutant. This reduced processivity is most easily detectable in the presence of competitor, which results in early stops in cDNA elongation (Fig. 5*b*).

The only difference between the wild-type and the N494D mutant enzyme detectable in this whole study was observed in the competition assay (Fig. 5). The hybrid-enzyme complex of this mutant is apparently sensitive to the competitor substrate before starting the polymerization reaction. This demonstrates that the mutation leads to a reduced stability of the hybrid-enzyme complex. It does not interfere with initial substrate binding or other enzyme functions.

Based on the x-ray crystallography studies of Davies *et al.* (1991), the amino acid N494 contributes to the tertiary conformation of the amino-terminal part of the protein. The residue N494 is not located at the active center of the retroviral RNase H (Davies *et al.*, 1991) and does not belong to the amino acids that form direct contacts to the substrate identified by Nakamura *et al.* (1991) for the RNase H of *E. coli*. N494 stabilizes an amino-terminal type 2 tight turn by interactions with A437 and I434 (Davies *et al.*, 1991). A disturbance of this interaction by site-directed mutagenesis might lead to an unfortunate conformation, resulting in a lower stability of the initial enzyme-substrate complex.

According to x-ray studies performed by Katayanagi *et al.* (1990, 1992) and Yang *et al.* (1990), the metal-binding site of the RNase H of *E. coli* is formed by the residues D10, E48, and D70. The substrate binding for this enzyme was characterized by ^{1}H-^{15}N heteronuclear two-dimensional NMR studies and confirmed by site-directed mutagenesis. Thereby amino acid N45, which corresponds to residue Q475 of the retroviral RNase H, has been identified as one of the residues that forms direct contact to the substrate. This result is in good agreement with the data presented here. A partially destroyed enzyme-substrate binding explains why the endonuclease and the subsequent 3′ → 5′ processive RNase H activity of the Q475E mutant are partially impaired (Fig. 2) and why the velocity of the RNase H activity is retarded (Fig. 4*a*). The enzyme contact to the substrate would therefore result in impaired RNA degradation in the presence of competitor (Fig. 5*a*). In addition a reduced processivity of the RNA-dependent DNA polymerase activity was observed, confirming the idea that the RT harbors only one substrate binding site, which is shared by the RNase H and the polymerase activity (Painter *et al.*, 1990; Krug and Berger, 1991). The impaired hydrolysis of homopolymeric substrates might be due to structural properties of homopolymeric- in contrast to heteropolymeric substrates.

Although the H539 residue does not contribute to the metal-binding or the substrate-binding amino acids, its localization is in close proximity to the active center and the substrate-binding site, suggesting a functional importance for both. Site-directed mutagenesis of this highly conserved residue led to the most highly impaired enzyme functions (Wöhrl *et al.*, 1991). They resemble those of the Q475E mutant enzyme described here.

Why are the residues Q475 and N494 so highly conserved in retroviral RNase H? When the Q475 is compared to the corresponding amino acid of the *E. coli* RNase H, only a conserved change is seen and may therefore exhibit the same function described for this amino acid in the *E. coli* RNase H, where it belongs to one of the substrate-binding amino acids. With N494 there is no corresponding residue in *E. coli*. Therefore this amino acid appears to be important only for retroviral RNase H. The retroviral enzymes are fused to a tether and polymerase domain and have therefore no free amino termini. This may require an additional stabilization of the amino-terminal conformation of the viral RNase H domain to guarantee accessibility of the hydrophobic pocket for substrate binding, which is not required for the *E. coli* RNase H. This may be guaranteed by the conserved N494 residue.

Acknowledgment—We are grateful to R. Oeltjen and Ian Henderson for excellent technical assistance. We also thank E. Philippi for preparing the drawings.

REFERENCES

Baltimore, D. (1970) *Nature* **226**, 1209–1211
Becerra, S. P., Clore, G. M., Gronenborn, A. M., Karlström, A. R., Stahl, S. J., Wilson, S. H., and Wingfield, P. T (1990) *FEBS Lett.* **270**, 76–80
Ben-Artzi, H., Zeelon, E., Gorecki, M., and Panet, A. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 927–931

Champoux, J. J., Gilboa, E., and Baltimore, D. (1984) *J. Virol.* **49**, 686–691
Crouch, R. J. (1990) *New Biol.* **9**, 771–777
Davies, J. F., Hostomska, Z., Hostomsky, Z., Jordan, S. R., and Matthews, D. A. (1991) *Science* **252**, 88–95
DeStefano, J. J., Buiser, R. G., Mallaber, L. M., Myers, T. W., Bambara, R. A., and Fay, P. J. (1991a) *J. Biol. Chem.* **266**, 7423–7431
DeStefano, J. J., Buiser, R. G., Mallaber, L. M., Bambara, R. A., and Fay, P. J. (1991b) *J. Biol. Chem.* **266**, 24295–24301
DiMarzo-Veronese, F., Copeland, T. D., and DeVico, A. L. (1986) *Science* **231**, 1289–1291
Doolittle, R. F., Feng, D.-F., Johnson, M. S., and McClure, M. A. (1989) *Q. Rev. Biol.* **64**, 1–29
Dudding, L. R., Harington, A., and Mizrahi, V. (1990) *Biochem. Biophys. Res. Commun.* **167**, 244–250
Dudding, L. R., Nkabinde, N. C., and Mizrahi, V. (1991) *Biochemistry* **30**, 10498–10506
Furfine, E. S., and Reardon, J. E. (1991) *J. Biol. Chem.* **266**, 406–412
Gilboa, E., Mitra, S. W., Goff, S., and Baltimore, D. (1979) *Cell* **18**, 93–100
Goff, S. P.(1990) *J. Acquired Immune Defic. Syndr.* **3**, 817–831
Hafkemeyer, P., Ferrari, E., Brecher, J., and Hübscher, U. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 5262–5266
Hansen, J., Schulze, T., and Moelling, K. (1987) *J. Biol. Chem.* **262**, 12393–12396
Hansen, J., Schulze, T., Mellert, W.and Moelling, K. (1988) *EMBO* **7**, 239–243
Hostomsky, Z., Hostomsky, Z., Hudson, G. O., Moomaw, E. W., and Nodes, B. R. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 1148–1152
Huang, S. C., Smith, J. R., and Moen, L. K. (1992) *Biochem. Biophys. Res. Commun.* **184**, 986–992
Huber, H. E., and Richardson, C. C. (1990) *J. Biol. Chem.* **265**, 10565–10573
Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T., and Morikawa, K. (1990) *Nature* **347**, 306–309
Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T. and Morikawa, K. (1992) *J. Mol. Biol.* **223**, 1029–1052
Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., and Steitz, T. A. (1992) *Science* **256**, 1783–1790
Krug, M. S., and Berger, S. L. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 3539–3543
Krug, M. S., and Berger, S. L. (1991) *Biochemistry* **30**, 10614–10623
Lei, M. H., and Verma, I. M. (1978) *J. Virol.* **25**, 652–663
Lori, F., Scovassi, A. I., Zella, D., Achilli, G., Cattaneo, E., Casoli, C., and Bertazzoni, U. (1988) *AIDS Res. Hum. Retroviruses* **4**, 393–398
Lowe, D. M., Aitken, A., Bradley, C., Darby, G. K., Larder, B. A., Powell, K. L., Purifoy, D. J., Tisdale, M., and Stammers, D. K. (1988) *Biochemistry* **27**, 8884–8889
Maniatis, T., Fritsch, E. F., and Sambrook, J.(1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Habor Laboratory, Cold Spring Harbor, NY
Mizrahi, V. (1989) *Biochemistry* **28**, 9088–9094
Mizrahi, V., Usdin, M. T., Harington, A., and Dudding, L. R.(1990) *Nucleic Acids Res.* **18**, 5359–5363
Moelling, K. (1992) *Med. Microbiol. Lett.* **1**, 71–77
Moelling, K., Bolognesi, D. P., Bauer, W., Büsen, W., Plassmann, H. W., and Hausen, P. (1971) *Nat. New Biol.* **234**, 240–243
Nakamura, H., Oda, Y., Iwai, S., Inoue, H., Ohtsuka, E., Kanaya, S., Kimura, S., Katsuda, C., Katayanagi, K., Morikawa, K., Miyashiro, H., and Ikehara, M. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 11535–11539
Omer, C. A., and Faras, A. J. (1982) *Cell* **30**, 797–805
Oyama, F., Kikuchi, R., Crouch, R. J., and Uchida, T. (1989) *J. Biol. Chem.* **264**, 18808–18817
Painter, G. R.,Wright, L. L., Hopkins, S., and Furman, P. A. (1990) *J. Cell. Biochem. Suppl.* **14D**, Abstr. L140 109
Prasad, V. R., and Goff, S. P. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 3104–3108
Resnick, R., Omer, C. A., and Faras, A. J. (1984) *J. Virol.* **51**, 813–821
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
Schatz, O., Cromme, F. V., Naas, T., Lindemann, D., Mous, J., and Le Grice, S. F. J. (1989) *Gene Regulation and AIDS* (Papas, T., ed) pp. 293–303, Portfolio Publishing, Houston, TX
Schatz, O., Mous, J., and Le Grice, S. F. J. (1990) *EMBO J.* **9**, 1171–1176
Starnes, M. C., and Chen, Y. (1989) *J. Biol. Chem.* **264**, 7073–7077
Tanese, N., and Goff, S. P. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 1777–1781
Temin, H. M., and Mitzutami, S. (1970) *Nature* **226**, 1211–1213
Tisdale, M., Larder, B. A., Lowe, D. M., Stamners, D. K., Purifoy, D. T. M., Emy, P., Bradley, C., Kemp, S., Darby, G., and Powell, K. L. (1989) *J. Antimicrobiol. Chemother.* **23**, Suppl. A47, S4
Tisdale, M., Schulze, T., Larder, B. A., and Moelling, K. (1991) *J. Gen. Virol.* **72**, 59–66
Varmus, H. (1987) *Sci. Am.* **257**, 56–64
Varmus, H., and Swanstrom, R. (1984) in *RNA Tumor Viruses* (Weiss, R., Teich, N., Varmus, H., and Coffin, J., eds) 2nd Ed., pp. 369–512, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Watson, K. F., Schendel, P. L., Rosok, M. J., and Ramsey, L. R. (1979) *Biochemistry* **18**, 3210–3219
Wintersberger, U. (1990) *Pharmacol. Ther.* **48**, 259–280
Wöhrl, B. M., and Moelling, K. (1990) *Biochemistry* **29**, 10141–10147
Wöhrl, B. M., Volkmann, S., and Moelling, K. (1991) *J. Mol. Biol.* **220**, 801–818
Yang, W., Hendrickson, W. A., Crouch, R. J., and Satow, Y. (1990) *Science* **249**, 1398–1405

February 5, 1993 VOLUME 268 NUMBER 4

ISSN 0021-9258
JBCHA3 268(4) 2269-3016 (1993)

THE Journal of Biological Chemistry

Published by the American Society for Biochemistry and Molecular Biology

FOUNDED BY CHRISTIAN A. HERTER

ENDOWED AND SUSTAINED IN PART BY THE CHRISTIAN A. HERTER MEMORIAL FUND

NCBI

Sequence Revision History

About Entrez

Search for Genes
LocusLink provides curated information for human, fruit fly, mouse, rat, and zebrafish

Entrez Nucleotide
Help | FAQ

Batch Entrez: Upload a file of GI or accession numbers to retrieve sequences

Check sequence revision history

How to create WWW links to Entrez

LinkOut

Cubby

Related resources
BLAST

Reference sequence project

Submit to GenBank

Revision history for "L10922"

GI	Version	Update Date	Status
407983	1	Nov 21 2002 11:26	Live
407983	1	Nov 21 2002 11:23	Dead
407983	1	Oct 4 1994 11:56	Dead
407983	1	Apr 19 1994 0:37	Dead
407983	1	Apr 9 1994 0:18	Dead
407983	1	Oct 15 1993 0:13	Dead

Accession L10922 was first seen at NCBI on Oct 15 1993 0:13

NCBI

Sequ nce Revision History

Revision history for "L10923"

GI	Version	Update Date	Status
407979	1	Oct 4 1994 11:56	Live
407979	1	Apr 19 1994 0:37	Dead
407979	1	Apr 9 1994 0:18	Dead
407979	1	Oct 15 1993 0:13	Dead

Accession L10923 was first seen at NCBI on Oct 15 1993 0:13

NCBI

Sequence Revision History

Revision history for "L10924"			
GI	**Version**	**Update Date**	**Status**
407987	1	Oct 4 1994 11:56	Live
407987	1	Apr 19 1994 0:37	Dead
407987	1	Apr 9 1994 0:18	Dead
407987	1	Oct 15 1993 0:13	Dead

Accession L10924 was first seen at NCBI on Oct 15 1993 0:13

NCBI

Sequ nc Revision History

About Entrez

Search for Genes
LocusLink provides curated information for human, fruit fly, mouse, rat, and zebrafish

Entrez Nucleotide
Help | FAQ

Batch Entrez: Upload a file of GI or accession numbers to retrieve sequences

Check sequence revision history

How to create WWW links to Entrez

LinkOut

Cubby

Related resources
BLAST

Reference sequence project

Submit to GenBank

Revision history for "J02342"

GI	Version	Update Date	Status
210171	1	Jun 7 1997 17:49	Live
210171	1	Mar 11 1996 15:07	Dead
210171	1	Mar 5 1996 0:40	Dead
210171	1	Oct 4 1994 7:19	Dead
210171	1	Apr 28 1993 0:40	Dead

Accession J02342 was first seen at NCBI on Apr 28 1993 0:40

JOURNAL OF VIROLOGY, Feb. 2001, p. 2002–2009
0022-538X/01/$04.00+0 DOI: 10.1128/JVI.75.4.2002–2009.2001

Vol. 75, No. 4

Evolution and Characterization of Tetraonine Endogenous Retrovirus: a New Virus Related to Avian Sarcoma and Leukosis Viruses

DEREK E. DIMCHEFF,* MALLIKA KRISHNAN, AND DAVID P. MINDELL

Department of Biology and Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48109-1079

Received 7 August 2000/Accepted 9 November 2000

In a previous study, we found avian sarcoma and leukosis virus (ASLV) *gag* genes in 19 species of birds in the order Galliformes including all grouse and ptarmigan (Tetraoninae) surveyed. Our data suggested that retroviruses had been transmitted horizontally among some host species. To further investigate these elements, we sequenced a replication-defective retrovirus, here named tetraonine endogenous retrovirus (TERV), from *Bonasa umbellus* (ruffed grouse). This is the first report of a complete, replication-defective ASLV provirus sequence from any bird other than the domestic chicken. We found a replication-defective proviral sequence consisting of putative Gag and Env proteins flanked by long terminal repeats. Reverse transcription-PCR analysis showed that retroviral *gag* sequences closely related to TERV are transcribed, supporting the hypothesis that TERV is an active endogenous retrovirus. Phylogenetic analyses suggest that TERV may have arisen via recombination between different retroviral lineages infecting birds. Southern blotting using *gag* probes showed that TERV occurs in tetraonines but not in chickens or ducks, suggesting that integration occurred after the earliest phasianid divergences but prior to the radiation of tetraonine birds.

Endogenous retroviruses have been found in all vertebrate hosts examined. These viruses are integrated into host genomes at multiple locations and are usually transmitted vertically via the germline. The most extensively studied endogenous avian retroviruses are found in the genome of the domestic chicken and belong to the avian sarcoma and leukosis virus (ASLV), or alpharetrovirus, genus. Endogenous ASLVs include Rous associated virus-0 (RAV-0), endogenous avian retroviruses (EAVs), and avian retrotransposons from chickens (ART-CH). RAV-0 is thought to represent a recently integrated endogenous virus, because its provirus DNA sequence is highly conserved relative to those of exogenous ASLVs such as Rous sarcoma virus (RSV). EAV-0 elements are thought to be older integrations of avian retroviruses, because they are less similar to exogenous viruses and their phylogeny closely reflects host phylogeny (5, 19). The number of known EAVs is increasing and includes recently identified EAV-HP and ev/J, which are apparently the same endogenous virus described independently (20, 21). ART-CH elements have deletions in all retroviral genes, but they retain *cis*-acting sequences necessary for retrotransposition (11, 17).

Recently, we showed that endogenous ASLVs are found in three families of galliform (fowl-like) birds, and, in some cases, the phylogenetic patterns observed for virus genes were incongruent with host phylogeny (8). Our findings are beginning to elucidate the ancient evolutionary association between retroviruses and birds, and they suggest the possibility of more-recent horizontal transmission of endogenous viruses between avian hosts as well.

In this report we describe a new avian proviral genome, obtained from a genomic library of *Bonasa umbellus* (ruffed grouse), called tetraonine endogenous retrovirus (TERV). Tetraonines are a subfamily of galliform birds consisting of grouse and ptarmigan. This is the first report of a complete, replication-defective ASLV provirus sequence from a bird other than the domestic chicken. We compare the structure of TERV to those of published avian retroviruses in order to investigate its function and evolution. Southern blot and reverse transcription-PCR (RT-PCR) analyses are used to document the distribution and expression of TERV-related viruses in galliform birds. We hypothesize that TERV is an active, endogenous retrovirus formed through recombination between endogenous retroviral lineages.

Generation of grouse λ bacteriophage genomic library and characterization of provirus structure. A *B. umbellus* lambda genomic library was constructed using a Lambda FIX II/*Xho*I partial-fill-in vector kit (Stratagene, La Jolla, Calif.). This library was screened by lifting plaques onto nylon membranes and probing with a ^{32}P-labeled *gag* probe. This probe was amplified by PCR using GAG.F1 and GAG.R1 primers (Fig. 1) and standard PCR conditions as previously described (8). After gel purification of PCR products, ≈25 ng of probe DNA was radiolabeled using [α-^{32}P]dATP (3,000 Ci/mmol; Amersham Pharmacia Biotech), 6 U of Klenow fragments, and random primers. Positive plaques were grown in liquid culture, and recombinant phage DNA was isolated using standard protocols (22). Provirus-positive phage DNA was randomly fragmented and subcloned into pZero vector (Stratagene). Plasmid DNA was isolated from positive colonies using a QIAprep spin miniprep kit (Qiagen) and sequenced using universal primers as previously described (8). The sequence was edited, and contigs were assembled, using Sequencher (Gene Codes Corp.). Published ASLV Sequences are listed with GenBank accession numbers in Table 1.

The DNA insert of one recombinant lambda was sequenced to reveal a replication-defective retroviral sequence of 3,711

* Corresponding author. Present address: Rocky Mountain Laboratories, 903 S. 4th St., Hamilton, MT 59840. Phone: (406) 363-9359. Fax: (406) 363-9286. E-mail: ddimcheff@nih.gov.

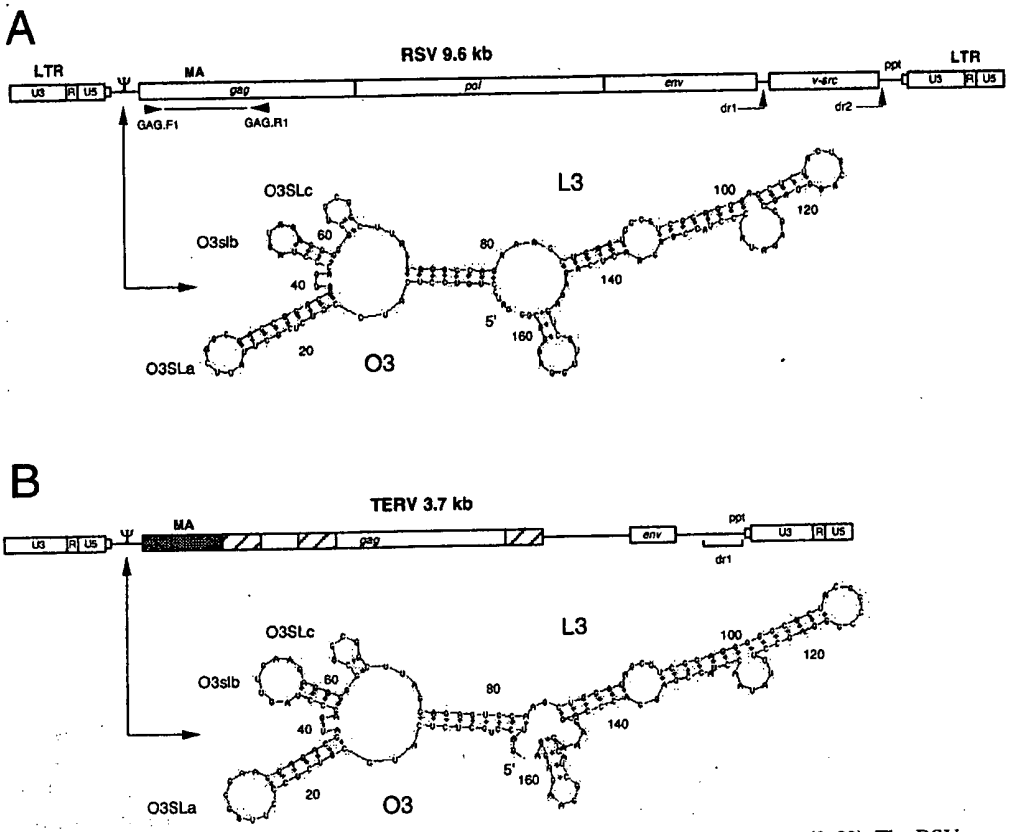


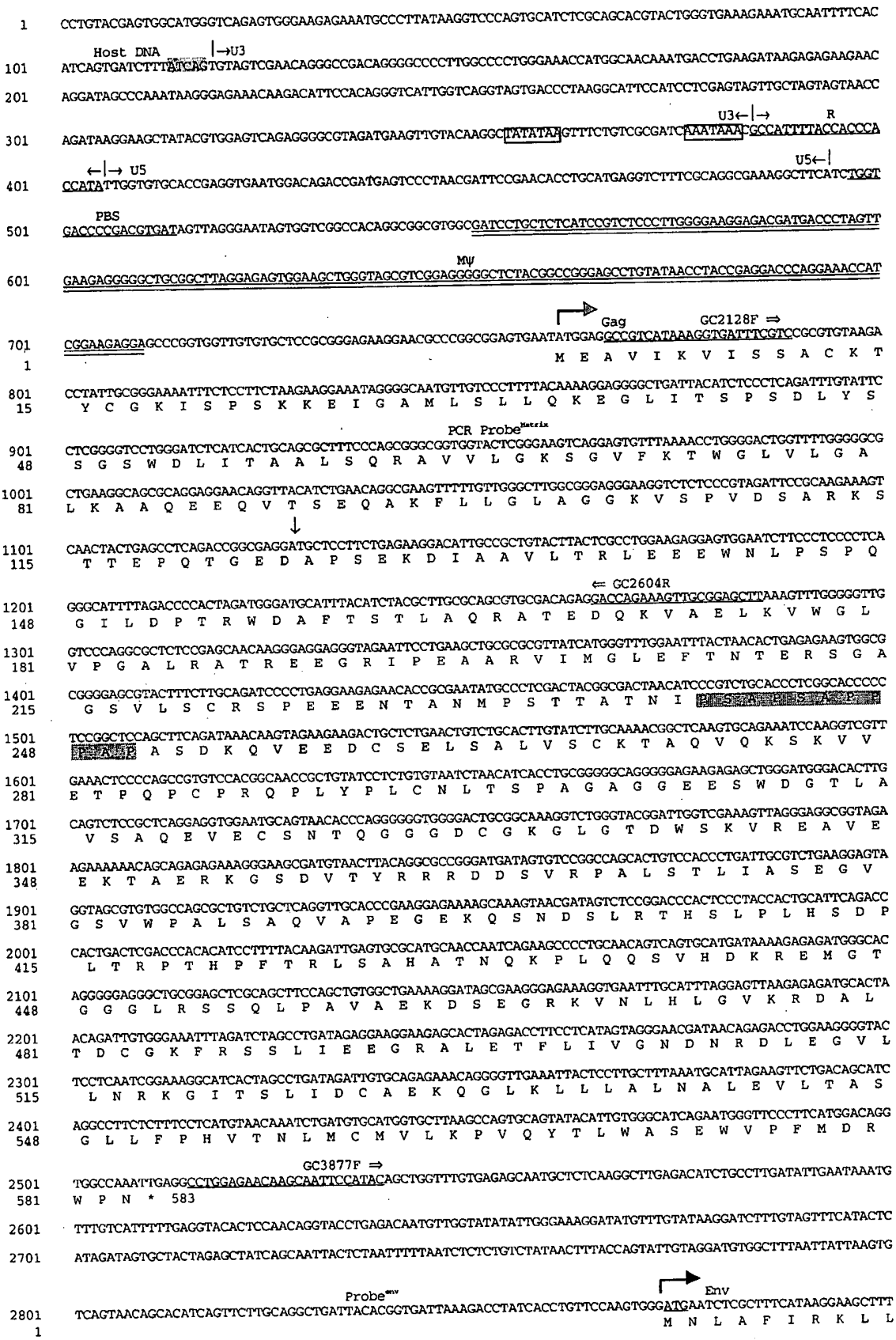
FIG. 1. Comparison of RSV (A) and TERV (B) genome and packaging signal (Ψ) secondary structures (3, 23). The RSV genome is not drawn to scale. The first 342 nucleotides (shaded region) of TERV *gag* are highly conserved relative to published ASLVs. The nucleotide similarity ranges from 80 (EAV-HP) to 97% (*B. umbellus* ASLV). The matrix (MA) region of Gag is indicated above coding regions. Hatched boxes, three regions in TERV with the greatest similarity (29 to 59%) to those of EAV-HP (20, 21). The single line between TERV *gag* and *env* denotes an apparent noncoding sequence. Primers GAG.F1 and GAG.R1 (8), used to amplify the probe for genomic library screens, are shown below RSV. Predicted secondary structures of the retroviral packaging sequence were modeled using Mfold (4, 32). Nucleotide positions listed are relative to the 5′ end of the packaging sequence (Fig. 2, MΨ). Major stem-loop structures are identified using the notation of Banks et al. (3). The RSV secondary structure is based on a consensus of 20 previously published ASLV packaging sequences. This structure is identical to the structure published by Banks and Linial (2).

bp, which we call TERV (Fig. 2). Host flanking sequences at the 5′ and 3′ ends of TERV (1,439 and 733 kb, respectively) were found to be not similar to any sequences in GenBank using BLAST (1). The TERV provirus genome structure was similar to those of other defective avian retroviruses such as Fujinami sarcoma virus, EAV-HP, and ART-CH (11, 20, 21, 24). The TERV protein coding sequence was flanked by long terminal repeats (LTRs), and these in turn were flanked by 5-bp direct repeats (5′-ATCAG-3′). The terminal nucleotides of the provirus were 5′-TG...CA-3′ and were part of imperfect indirect repeats (Fig. 2).

We compared TERV untranslated region (UTR) sequences to retrovirus UTR sequences of known function to identify potentially functionally homologous regions. The 3′ boundary of the 5′ LTR was distinguished by a sequence highly conserved relative to those of other avian retrovirus $tRNA^{Trp}$ primer binding sites (PBS) (initiation of minus-strand synthesis). The 5′ boundary of the 3′ LTR was distinguished by the polypurine tract sequence (PPT) (initiation of plus-strand synthesis). Based on these boundaries, host flanking direct repeat sequences, and sequence conservation with other ASLV R

TABLE 1. Virus names and GenBank accession numbers

Virus	Accession no.
Gallus ASLVs	
RSV	D10652
RAV-0	M73497, J02295, and J02015
Avian sarcoma virus Y73	J02027
Avian myeloblastosis virus	L10922
Fujinami sarcoma virus	AF033810
EAV-HP	J238125 and AJ238124
EAV-0	M31065
ART-CH	L25262
Gallus varius ASLV	AF225361
Tetraonine/odontophorid ASLVs	
TERV	AF289082
B. umbellus ASLV	AF225316
Colinus virginianus ASLV	AF225336
Lagopus lagopus ASLV	AF225366
Phasianus colchicus ASLV	AF225384
Mammalian retroviruses	
Porcine endogenous virus	Y12239
Murine leukemia virus	AF064089


Host DNA
U3
R
U5
PBS
MΨ
Gag
GC2128F
PCR Probe
GC2604R
GC3877F
Probe
Env

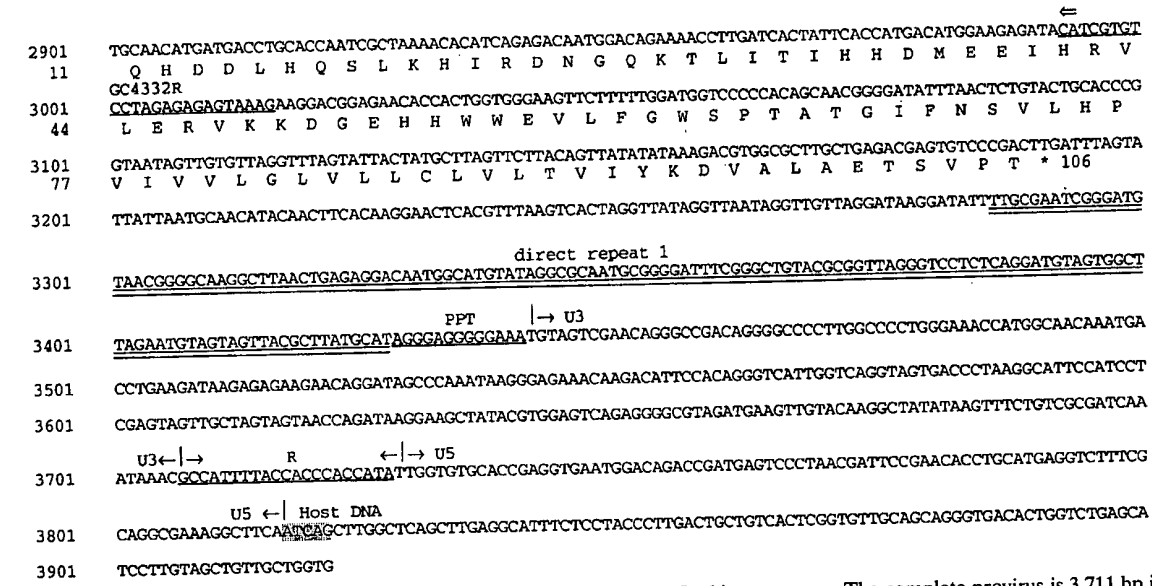


FIG. 2. Complete nucleotide sequence of a TERV provirus including the host flanking sequence. The complete provirus is 3,711 bp in length. Direct repeats (5 bp) flanking the provirus are shaded, and major LTR regions are indicated. The repeat region (R), $tRNA^{Trp}$ PBS, and PPT are underlined. The TATA box and polyadenylation signal are boxed. Translations of Gag and Env proteins are below the nucleotide sequence. Downward arrow, junction between the highly conserved *gag* sequence (5′) and regions of low conservation (3′). The shaded region in Gag corresponds to a region possibly homologous to the L domain (PPPPY) in ASLVs. Double underlining, putative packaging sequences (MΨ and direct repeat 1). Only one copy of RSV-related direct repeat 1 was found in TERV. Primer sequences described in the text are underlined, and the primer names are indicated. $Probe^{Matrix}$ and $Probe^{env}$, fragments representing probes used for Southern hybridization.

regions, identical LTRs of 376 bp were defined. The typical proviral LTR structure (5′-U3-R-U5-3′) was found, and lengths for the U3, R, and U5 regions were 266, 21, and 89 bp, respectively. These LTRs were longer than LTRs from RSV (327 bp) (23), EAV-HP (315 bp) (20, 21), EAV-0 (243 bp) (5, 6), and RAV-0 (277) (13) but slightly shorter than ART-CH LTRs (388 bp) (17).

The TERV U3 sequence was most similar to that of RAV-0, particularly at conserved motifs that function in retrovirus transcription. Putative transcription factors were identified by searching on-line database TRANSFAC, specifying a threshold score of 85 (12), and by comparative sequence analysis with known transcriptional regulatory sequences from ASLVs. The 5′-most enhancer binding element was similar to an IK-2 element (position −221), which is bound by a transcription factor isolated from mice (10). The next sequence, a δEFI binding element, was located at −194 in relation to the transcription start site. It is thought that protein δEFI is an embryonic gene regulator in chickens (9). These two elements were not known from other avian retrovirus LTRs; thus further characterization is necessary as these elements show limited similarity to the consensus sequences. At position −165 there was a putative serum response element (SRE) (27, 28) that was highly conserved relative to the SRE found in RAV-0 (31). Within this element there was a CArG box, defined as 5′-CC(A/T)$_6$GG-3′. This sequence was followed by an inverted CCAAT box at −132 also found in RAV-0 and RSV. The CCAAT box was in the middle of putative NF-Y and C/EBP binding sites. Finally, another SRE was located at −75.

The U3 region of TERV was highly conserved relative to those of other ASLVs, starting with the TATA box located at position −24, followed by a polyadenylation signal just upstream of the repeat region (R). The R and U5 regions of TERV had the greatest sequence similarity to those of RAV-0 at 91 and 89%, respectively, and had 82 and 86% similarity, respectively, to those of RSV (Table 2). The $tRNA^{Trp}$ PBS were identical across all avian retroviruses mentioned in this report with the exception of that of EAV-0, which had an additional cytosine residue (Table 2).

Analysis of putative Gag and Env proteins. Two open reading frames (ORFs) that had sequence similarity to those of published ASLVs were found. The N-terminal 123 amino acids encoded by the first ORF had 88% identity to RSV and 95 to

TABLE 2. Percentages of DNA sequence similarity among UTRs of TERV and other avian retroviruses

Virus[b]	% Similarity[a] to TERV region:							
	U3	TATA box	TATA to R	R	U5	PBS	MΨ[c]	dr1[d]
RSV	37.42	85.71	67.74	81.82	85.57	100.00	**80.40**	**88.65**
FuSV	42.11	**100.0**	67.74	82.07	87.88	100.00	80.27	57.85
AMV	35.18	71.42	48.39	78.26	80.43	100.00	73.78	86.58
Y73	42.12	85.71	64.52	86.36	81.59	100.00	79.59	88.49
RAV-0	**67.86**[b]	**100.0**	**77.42**	**90.91**	**89.03**	100.00	ND[e]	88.62
ART-CH	32.44	**100.0**	65.71	62.73	44.07	100.00	31.35	44.24
EAV-HP	35.22	85.71	45.16	62.73	44.06	100.00	33.68	47.78
EAV-0	32.59	**100.0**	69.58	**90.91**	66.30	100.00	ND	ND

[a] Percent similarities were calculated using uncorrected *P* distances ([1 − *P*] 100). Boldface numbers indicate the greatest similarity to TERV.

[b] FuSV, Fujinami sarcoma virus; AMV, avian myeloblastosis virus; Y73, avian sarcoma virus Y73.

[c] MΨ comparisons for RSV, FuSV, AMV, and Y73 are based on the alignment presented in Banks and Linial (2).

[d] The dr1 sequence (117 bp) is located adjacent to the 5′ end of the PPT.

[e] ND, no data were available for comparison.

97% identity to previously published *B. umbellus* Gag sequences corresponding to the M domain of the matrix region (Fig. 1). The remaining 460 amino acids had 29 to 59% identity with those of ASLV Gag proteins in three regions, each separated by sequences that showed no significant sequence similarity to published ASLV amino acid sequences in GenBank (Fig. 1). The regions of EAV-HP had the highest similarity to these regions, slightly higher than published regions of *B. umbellus* ASLV. One other region of amino acid similarity to Gag begins at P^{240} and is a proline-rich region (PSAPSAPPPAP) possibly homologous to the L domain (PPPPY) found in all ASLVs, which functions in virus assembly (30).

A second ORF, located 360 bp downstream of the *gag* gene, encoded 106 amino acids that had 29 to 40% identity to Gag of ASLV and murine leukemia virus and porcine endogenous retrovirus Env based on BLAST search results (Fig. 1). The putative ORF corresponded to the carboxyl terminus of the transmembrane region of Env. This Env-related ORF was just upstream of a 110-bp sequence with 89% identity to those of exogenous ASLVs, which corresponds to the direct repeat 1 sequence found in all ASLV genomes studied, including replication-defective transforming avian retroviruses.

Analysis of *gag* gene transcription. Sequence analysis suggests that TERV is capable of transcription. To determine if *gag* sequences were transcribed, total RNA was extracted from heart muscle of one adult *B. umbellus* animal and whole 8-day-old *Phasianus colchicus* and *Colinus virginianus* embryos using Trizol reagent (Gibco Life Technologies) according to the manufacturer's protocol. RNA extract was treated with 5 U of DNase I, amplification grade (Gibco Life Technologies), to eliminate DNA contamination according to the manufacturer's protocol. RT-PCR was performed on DNase-treated total RNA using a Titan one-tube RT-PCR kit (Roche) and two sets of primers at an annealing temperature of 55°C. Primers GAG.F1 and GAG.R1 (8) are general *gag* primers, while GC2128F and GC2604R (Fig. 2) are specific to TERV *gag*. Chicken β-actin primers were used as a positive control for RNA. These primers (β-actinF [5′-AATGAGAGGTTCAGG TGCCC-3′] and β-actinR [5′-ATCACAGGGGTGTGGGTG TT-3′]) amplify a 410-bp fragment. RT-PCR products were verified by DNA sequence analysis as described previously (8). PCR using GAG.F1 and GAG.R1 primers was performed on DNase-treated and untreated samples to verify that DNA contamination was eliminated.

RT-PCR using GAG.F1 and GAG.R1 primers on DNase-treated samples resulted in products from all three birds as did PCR analysis of non-DNase-treated samples (Fig. 3). Size variation in amplicons is consistent with results obtained in our previous study of avian retroviral *gag* sequences (8). Sequence analysis of the RT-PCR product from *Phasianus* suggested that these primers amplified nontarget transcripts, whereas sequence analysis confirmed that *gag* transcripts were amplified from *Bonasa* and *Colinus*. RT-PCR using TERV-specific primers (GC2128F and GC2604R) resulted in amplification in all three birds. The product from *B. umbellus* was 475 bp, the size predicted from TERV, while the predominant products from *Phasianus* and *Colinus* were 200 to 300 bp larger (Fig. 3). These RT-PCR products were sequenced, and all were verified as *gag* sequences.

Normally, following transcription, two copies of the retroviral RNA genome are incorporated into viral particles. This packaging process involves the recognition and binding of sequence Ψ on the RNA genome by viral proteins. In ASLVs, a 160-nucleotide sequence has been identified as the minimal packaging signal (MΨ) in the leader region between the PBS and the Gag initiation codon (3). TERV contained a sequence highly conserved relative to avian retrovirus MΨs located between the primer binding site and the *gag* initiation codon. TERV MΨ had a similarity of 78.2% to a consensus alignment (2) of 20 exogenous and endogenous avian retrovirus packaging sequences. EAV-HP and ART-CH did not have the same level of conservation in the packaging signal. The sequences between the PBS and the *gag* initiation codon in ART-CH and EAV-HP are about 100 bp shorter than those in TERV and other ASLVs.

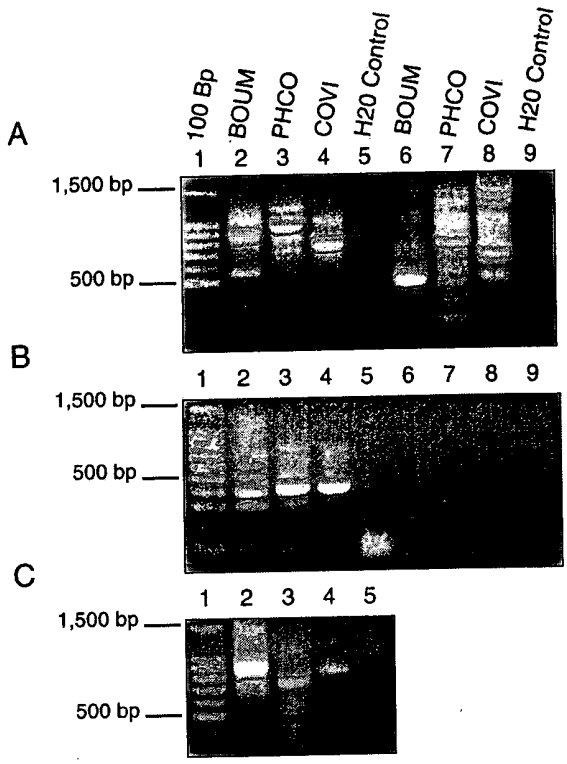


FIG. 3. RT-PCR analysis of total RNA isolated from heart muscle for an adult *B. umbellus* animal (BOUM) and 8-day-old embryos of *P. colchicus* (PHCO) and *C. virginianus* (COVI). (A) RT-PCR on total RNA. Lane 1, 100-bp ladder (Promega); lanes 2 to 5, primers GAG.F1 and GAG.R1; lanes 6 to 9, primers GC2128F and GC2604R. (B) RT-PCR using β-actin primers (lanes 2 to 5) and PCR using primers GAG.F1 and GAG.R1 on DNase-treated RNA extracts (lanes 6 to 9). (C) PCR on non-DNase-treated samples using primers GAG.F1 and GAG.R1.

It appears that the secondary structure of MΨ plays a significant role in efficient packaging (2). TERV and the ASLV consensus packaging sequence were analyzed using the Mfold program (version 3.0) to model the most-stable secondary structures (4, 16). The lowest free energies were −61.02 kcal/mol for the secondary structure of TERV and −56.82 kcal/mol for the folding of a consensus sequence of 20 ASLV packaging signals (Fig. 1). The two modeled structures were remarkably

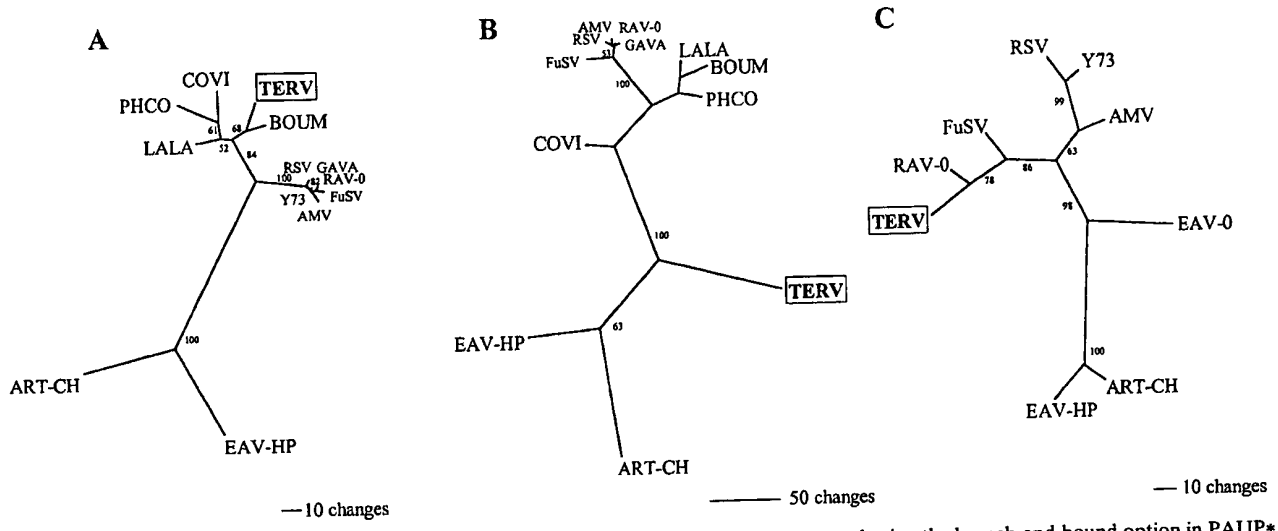


FIG. 4. Phylogenetic analyses of *gag* and UTR sequences. Unrooted MP trees were constructed using the branch-and-bound option in PAUP* (25). Bootstrap values along branches were calculated using 100 random-addition replicate searches. COVI, *C. virginianus* ASLV; PHCO, *P. colchicus* ASLV; LALA, *Lagopus lagopus* ASLV; BOUM, *B. umbellus* ASLV; GAVA, *G. varius* ASLV; FuSV, Fujinami sarcoma virus; AMV, avian myeloblastosis virus; Y73, avian sarcoma virus Y73. Branch lengths are proportional to inferred amounts of evolutionary change. (A) Relationship of 10 previously published ASLV and TERV sequences based on 342 nucleotide sites from the matrix region of the *gag* gene. (B) One of two equally parsimonious trees for the relationship of 11 published ASLVs and TERV based on the more-divergent amino acid region of Gag downstream from the region analyzed in panel A. The two equally parsimonious trees differ only in the placement of GAVA. (C) Phylogeny of eight previously published UTR sequences and TERV.

similar. TERV MΨ had the two major stem-loop regions O3 and L3 and the three minor stem-loops O3SLa, O3SLb, and O3SLc previously identified by Banks and Linial (2). In ART-CH and EAV-HP the most-stable secondary structures for sequences that correspond to MΨ were not similar to those of RSV, the consensus of ASLVs, or TERV when folded with Mfold (not shown).

Our previous work showing that some ASLV phylogenetic relationships reflect overlapping host species ranges rather than host species phylogeny (8) suggests that horizontal transmission of ASLVs has occurred in the past. Here we have demonstrated that TERV is transcribed and contains sequences required for packaging and retrotransposition. We have no direct evidence regarding the means by which TERV or other tetraonine ASLVs move between host species; however, one possible mechanism is hybridization among host species, which is known to occur naturally between some tetraonines (14). Alternatively, TERVs could have been transmitted horizontally if they were transcribed and packaged with replication-competent retroviruses.

Phylogeny and recombination. To explore the relationship of TERV and other endogenous and exogenous ASLVs, we conducted phylogenetic analyses using three regions of TERV. Sequence alignments corresponding to various regions of TERV and ASLVs were performed using Clustal X (26). Phylogenetic analyses were performed using maximum parsimony (MP) as implemented in PAUP* (25). Branch-and-bound MP analyses were conducted, and bootstrap values were determined using 100 replicate searches.

The first region examined corresponds to the matrix gene of *gag*. MP analysis using this region (342 nucleotides) yielded one tree composed of three major groups (Fig. 4A). The first group consisted of endogenous and exogenous viruses isolated from birds in the genus *Gallus*, including the domestic chicken (*Gallus gallus*). The second group consisted of presumably endogenous proviruses from grouse and ptarmigan (Tetraoninae), and the third group consisted of endogenous viruses EAV-HP and ART-CH. TERV was the sister taxon to ASLV, whose *gag* gene was previously sequenced from *B. umbellus*.

The second region corresponded to the remaining sequence in TERV Gag adjacent to the highly conserved matrix region. Only short stretches of amino acids could be aligned to other ASLV Gag proteins, resulting in a data set with 125 parsimony-informative characters. Two equally parsimonious trees that differed only in the placement of *Gallus varius* ASLV were found (Fig. 4B). Interestingly, TERV and the other *B. umbellus* ASLV sequences were not sister taxa based on this portion of the *gag* gene. Instead, TERV is located between an EAV-HP/ART-CH group and a group containing the remaining ASLVs.

The third region analyzed was aligned with eight published avian retrovirus UTR sequences. This phylogeny (Fig. 4C) showed relationships similar to those from the tree shown in Fig. 4A. We found that EAV-HP and ART-CH formed a group that was sister to EAV-0 (EAV-0 was not included in *gag* phylogeny). TERV was most closely related to endogenous avian retrovirus RAV-0. No UTR sequences from other tetraonine retroviruses were available for comparison.

Recombination within the *gag* gene of RSV can occur with a relatively high frequency (15). Incongruent trees from our phylogenetic analyses suggest the possibility that TERV was formed by recombination between retroviruses. One recombination point may occur near amino acid 123, where the similarity of TERV Gag to *B. umbellus* Gag drops drastically from 97 to around 40%. Phylogenetic analysis illustrates that this downstream region of TERV Gag is not sister to those of other

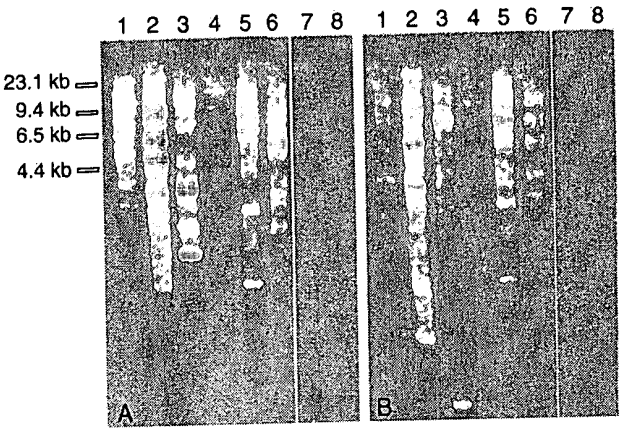


FIG. 5. Southern blot analysis of *Hin*dIII-digested avian genomic DNA using TERV probe^{matrix} and probe^{env}. Hybridization was performed at 65°C, and the filter was washed with salt-SDS solution at concentrations as low as 0.5× SSPE–0.5% SDS at 65°C. (A) Probe^{matrix}. (B) Probe^{env}. Lanes 1 and 2, *B. umbellus* (ruffed grouse); lanes 3, *Bonasa bonasia* (hazel grouse); lanes 4, *Lagopus lagopus* (willow ptarmigan); lanes 5, *Lagopus leucurus* (white-tailed ptarmigan); lanes 6, *Dendragapus falcipennis* (Siberian grouse); lanes 7, Anseriformes (*Aythya americana* [redhead]); lanes 8, *G. gallus* (domestic chicken). This shows the presence of TERV or closely related elements in all tetraonine birds surveyed.

tetraonine ASLVs, RAV-0, or exogenous ASLVs (Fig. 4B). Phylogenetic analysis and amino acid identity of a short region of Env (transmembrane region) suggest a relationship that is also incongruent with UTR and 5′ *gag* phylogenies (not shown). These findings are compatible with the interpretation that TERV was formed by recombination, although the parental sequences for divergent regions of TERV have yet to be discovered. An alternative explanation to recombination is that highly divergent regions of TERV result from differing selective pressures on the viral genome. We are examining additional complete tetraonine retrovirus genomes to investigate this possibility.

Distribution of TERV-related sequences. To determine the distribution of TERV-related sequences, genomic DNA was isolated from six galliform species and one anseriform (*Aythya americana*) using standard protocols. Roughly 3 μg of genomic DNA was digested to completion with *Hin*dIII, electrophoresed in a 1.0% agarose gel, and blotted overnight onto a positively charged nylon membrane (Hybond-XL; Amersham Pharmacia Biotech). Two different probes, probe^{matrix} and probe^{env}, were amplified by PCR from TERV-lambda DNA. These probes corresponded to a region of matrix (GC2128F and GC2604R) and a divergent region similar to that in which *env* is located (GC3877F and GC4332R) (Fig. 2). Hybridization was performed at 65°C in standard buffer overnight with probe^{matrix} first, followed by probe^{env}. The nylon membrane was washed in several steps with decreasing salt and sodium dodecyl sulfate (SDS) concentrations, down to 0.5× SSPE (1× SSPE is 0.18 M NaCl, 10 mM NaH_2PO_4, and 1 mM EDTA [pH 7.7]) and 0.1% SDS, and increasing temperature to 65°C. Prior to being probed with probe^{env}, the membrane was stripped by washing with 0.1% SDS at 100°C for 30 min and exposed to film for 1 week to verify that all of probe^{matrix} was removed.

Results of Southern analysis at high stringency using both probes were positive for the five grouse and ptarmigan tested and negative for a domestic chicken and a redhead (Fig. 5). Genomic DNA from two *B. umbellus* individuals was analyzed and had almost identical banding patterns. Both probes showed complex hybridization patterns consistent with the idea that TERV integrated in multiple locations within the bird genome.

Divergence date estimates from fossils suggest that tetraonines separated from their putative closest extant relative (*Meleagris gallopavo* [turkey]) in the mid-Miocene (15 to 20 million years ago), while modern tetraonines seem to have been present for at least 1 million years (7, 29). If our limited sampling of birds is representative, TERV could have integrated into the genome of a tetraonine sometime during the past 15 million years. Future surveys, including more Galliformes as well as birds from additional avian orders, are needed for more-reliable estimates of TERV age and relative timing of integration into host species genomes. The 100% identity between TERV LTRs suggests that TERV was active quite recently, although this activity may have been restricted within the genome.

Nucleotide sequence accession number. The TERV sequence obtained in this study has been assigned GenBank accession no. AF289082.

This work was supported by University of Michigan graduate student block grant funds to D.E.D. and M.K.

We thank Michael Frohlich, David Parker, and Vici Blanc for excellent technical assistance and Sergei Drovetski of the University of Washington Burke Museum for providing some tissue samples used in this study.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.
2. **Banks, J. D., and M. L. Linial.** 2000. Secondary structure analysis of a minimal avian leukosis-sarcoma virus packaging signal. J. Virol. **74**:456–464.
3. **Banks, J. D., A. Yeo, K. Green, F. Cepeda, and M. L. Linial.** 1998. A minimal avian retroviral packaging sequence has a complex structure. J. Virol. **72**: 6190–6194.
4. **Barciszewski, J., and B. F. C. Clark (ed.).** 1999. RNA biochemistry and biotechnology. Kluwer, Dordrecht, The Netherlands.
5. **Boyce-Jacino, M. T., K. O'Donoghue, and A. J. Faras.** 1992. Multiple complex families of endogenous retroviruses are highly conserved in the genus *Gallus*. J. Virol. **66**:4919– 4929.
6. **Boyce-Jacino, M. T., R. Resnick, and A. J. Faras.** 1989. Structural and functional characterization of the unusually short long terminal repeats and their adjacent regions of a novel endogenous avian retrovirus. Virology **173**:157–166.
7. **del Hoyo, J., A. Elliott, J. Sargatal, and J. Cabot.** 1992. Handbook of the birds of the world. New World vultures to guineafowl, vol. 2. Lynx Edicions, Barcelona, Spain.
8. **Dimcheff, D. E., S. V. Drovetski, M. Krishnan, and D. P. Mindell.** 2000. Cospeciation and horizontal transmission of avian sarcoma and leukosis virus *gag* genes in galliform birds. J. Virol. **74**:3984–3995.
9. **Funahashi, J., R. Sekido, K. Murai, Y. Kamachi, and H. Kondoh.** 1993. Delta-crystallin enhancer binding protein delta EF1 is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis. Development **119**:433–446.
10. **Georgopoulos, K., D. D. Moore, and B. Derfler.** 1992. Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. Science **258**:808– 812.
11. **Gudkov, A. V., E. A. Komarova, M. A. Nikiforov, and T. E. Zaitsevskaya.** 1992. ART-CH, a new chicken retroviruslike element. J. Virol. **66**:1726–1736.
12. **Heinemeyer, T., X. Chen, H. Karas, A. E. Kel, O. V. Kel, I. Liebich, T. Meinhardt, I. Reuter, F. Schacherer, and E. Wingender.** 1999. Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms. Nucleic Acids Res. **27**:318–322.
13. **Hughes, S. H.** 1982. Sequence of the long terminal repeat and adjacent

segments of the endogenous avian virus Rous-associated virus O. J. Virol. 43:191–200.
14. **Johnsgard, P. A.** 1983. The grouse of the world. University of Nebraska Press, Lincoln.
15. **Linial, M., and S. Brown.** 1979. High-frequency recombination within the *gag* gene of Rous sarcoma virus. J. Virol. 31:257–260.
16. **Mathews, D. H., J. Sabina, M. Zuker, and D. H. Turner.** 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288:911–940.
17. **Nikiforov, M. A., and A. V. Gudkov.** 1994. ART-CH: a VL30 in chickens? J. Virol. 68:846–853.
18. **Oh, W. K., J. C. Yoo, D. Jo, Y. H. Song, M. G. Kim, and D. Park.** 1997. Cloning of a SH3 domain-containing proline-rich protein, p85SPR, and its localization in focal adhesion. Biochem. Biophys. Res. Commun. 235:794–798.
19. **Resnick, R. M., M. T. Boyce-Jacino, Q. Fu, and A. J. Faras.** 1990. Phylogenetic distribution of the novel avian endogenous provirus family EAV-0. J. Virol. 64:4640–4653.
20. **Ruis, B. L., S. J. Benson, and K. F. Conklin.** 1999. Genome structure and expression of the ev/J family of avian endogenous viruses. J. Virol. 73:5345–5355.
21. **Sacco, M. A., D. M. Flannery, K. Howes, and K. Venugopal.** 2000. Avian endogenous retrovirus EAV-HP shares regions of identity with avian leukosis virus subgroup J and the avian retrotransposon ART-CH. J. Virol. 74: 1296–1306.
22. **Sambrook, J., T. Maniatis, and E. F. Fritsch.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell 32:853–869.
24. **Shibuya, M., and H. Hanafusa.** 1982. Nucleotide sequence of Fujinami sarcoma virus: evolutionary relationship of its transforming gene with transforming genes of other sarcoma viruses. Cell 30:787–795.
25. **Swofford, D. L.** 1999. PAUP*: phylogenetic analysis using parsimony (*and other methods), 4.0 ed. Sinauer, Sunderland, Mass.
26. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–4882.
27. **Treisman, R.** 1986. Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factors. Cell 46:567–574.
28. **Treisman, R.** 1985. Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5′ element and c-fos 3′ sequences. Cell 42:889–902.
29. **Unwin, D. M.** 1993. The fossil record 2, p. 717–737. *In* M. J. Benton (ed.), Aves. Chapman and Hall, New York, N.Y.
30. **Wills, J. W., C. E. Cameron, C. B. Wilson, Y. Xiang, R. P. Bennett, and J. Leis.** 1994. An assembly domain of the Rous sarcoma virus Gag protein required late in budding. J. Virol. 68:6605–6618.
31. **Zachow, K. R., and K. F. Conklin.** 1992. CArG, CCAAT, and CCAAT-like protein binding sites in avian retrovirus long terminal repeat enhancers. J. Virol. 66:1959–1970.
32. **Zuker, M., D. H. Mathews, and D. H. Turner.** 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide, p. 11–43. *In* J. Barciszewski and B. F. C. Clark (ed.), RNA biochemistry and biotechnology. Kluwer, Dordrecht, The Netherlands.

ICTVdB
Index of Viruses

Nucleic Acid:

General Host:

Family 00.061. *Retroviridae*

Taxonomic Structure of the Family

Family	00.061. *Retroviridae*
Genus	00.061.0.03. *Alpharetrovirus*
Genus	00.061.0.01. *Betaretrovirus*
Genus	00.061.0.02. *Gammaretrovirus*
Genus	00.061.0.05. *Deltaretrovirus*
Genus	00.061.0.08. *Epsilonretrovirus*
Genus	00.061.0.06. *Lentivirus*
Genus	00.061.0.07. *Spumavirus*

Genus 00.061.0.03. *Alpharetrovirus*

Type Species 00.061.0.03.001. *Avian leukosis virus* (ALV)

List of Species in the Genus

The ICTVdB virus code and the viruses. Official virus *species names* are in italics. Tentative virus s names ( ), isolates, strains, serotypes, or subspecies are not italicized.

Virus codes, virus names, arthropod vector and host names { }, serotypes, genome sequence access abbreviations ( ), are:

Species, their serotypes, strains and isolates

00.061.0.03.003.	*Avian leukosis virus*	
00.061.0.03.003.00.001.	Avian leukosis virus - RSA	[M37980]
00.061.0.03.003.00.001.	Avian leukosis virus - HPRS103	[Z46390]
	Oncogene containg viruses:	
	Replication competent viruses:	
00.061.0.03.007.	*Rous sarcoma virus*	
00.061.0.03.007.00.001.	Rous sarcoma virus (Prague C)	[J02342]
00.061.0.03.007.00.002.	Rous sarcoma virus (Schmidt-Ruppin B)	[AF052428
00.061.0.03.007.00.003.	Rous sarcoma virus (Schmidt-Ruppin D)	[D10652]
	Replication defective viruses:	

00.061.0.03.002.	*Avian carcinoma Mill Hill virus 2*	[K02082]
00.061.0.03.004.	*Avian myeloblastosis virus*	[J02013]
00.061.0.03.005.	*Avian myelocytomatosis virus 29*	[J02019]
00.061.0.03.010.	*Avian sarcoma virus CT10*	[Y00302]
00.061.0.03.006.	*Fujinami sarcoma virus*	[J02194]
00.061.0.03.008.	*UR2 sarcoma virus*	[M10455]
00.061.0.03.008.	(University of Rochester virus 2)	
00.061.0.03.008.	(Avian sarcoma virus UR-2)	
00.061.0.03.009.	*Y73 sarcoma virus*	[J02027]

Tenative Species in the Genus

None reported.

Genus 00.061.0.01. *Betaretrovirus*

Type Species 00.061.0.01.001. *Mouse mammary tumor virus* (MMTV)

List of Species in the Genus

The ICTVdB virus code and the viruses. Official virus *species names* are in italics. Tentative virus s names ( ), isolates, strains, serotypes, or subspecies are not italicized.

Virus codes, virus names, arthropod vector and host names { }, serotypes, genome sequence access abbreviations ( ), are:

Species, their serotypes, strains and isolates

00.061.0.01.002.	*Langur virus*	
00.061.0.01.003.	*Mason-Pfizer monkey virus*	[M12349]
00.061.0.01.003.00.001.	Simian retrovirus 1	[M11841]
00.061.0.01.003.00.002.	Simian retrovirus 2	[M16605]
00.061.0.01.004.	*Mouse mammary tumor virus*	[M15122]
00.061.0.01.005.	*Ovine pulmonary adenocarcinoma virus*	
00.061.0.01.005.	(Jaagsiekte sheep retrovirus)	[M80216]
00.061.0.01.006.	*Squirrel monkey retrovirus*	[M23385]

Tentative Species in the Genus

None reported.

Genus 00.061.0.02. *Gammaretrovirus*

Type Species 00.061.0.02.001. *Murine leukemia virus* (MLV)

List of Species in the Genus

The ICTVdB virus code and the viruses. Official virus *species names* are in italics. Tentative virus s names ( ), isolates, strains, serotypes, or subspecies are not italicized.

Virus codes, virus names, arthropod vector and host names { }, serotypes, genome sequence access abbreviations ( ), are:

Species, their serotypes, strains and isolates

Mammalian virus group:

Replication competent viruses:

00.061.0.02.004.	*Feline leukemia virus*	[M18247]
00.061.0.02.008.	*Gibbon ape leukemia virus*	[M26927]
00.061.0.02.009.	*Guinea pig type C oncovirus*	
00.061.0.02.014.	*Murine leukemia virus*	
00.061.0.02.014.01.001.	Abelson murine leukemia virus	[J02009]
00.061.0.02.014.01.002.	AKR (endogenous) murine leukemia	[J01998]
00.061.0.02.014.01.003.	Friend murine leukemia virus	
00.061.0.02.014.01.003.001.	FrMLV, isolate PVC211	[M93134]
00.061.0.02.014.01.003.002.	FrMLV, isolate FB29	[Z11128]
00.061.0.02.014.01.004.	Moloney murine leukemia virus	[J02255]

Replication defective viruses:

00.061.0.02.015.	*Porcine type C oncovirus*	
00.061.0.02.005.	*Finkel-Biskis-Jinkins murine sarcoma*	[K02712]
00.061.0.02.007.	*Gardner-Arnstein feline sarcoma virus*	
00.061.0.02.010.	*Hardy-Zuckerman feline sarcoma virus*	
00.061.0.02.011.	*Harvey murine sarcoma virus*	
00.061.0.02.012.	*Kirsten murine sarcoma virus*	
00.061.0.02.013.	*Moloney murine sarcoma virus*	[J02266]
00.061.0.02.016.	*Snyder-Theilen feline sarcoma virus*	
00.061.0.02.017.	*Woolly monkey sarcoma virus*	[J02394]
00.061.0.02.017.	(Simian sarcoma virus)	

Reptilian virus group:

00.061.0.02.018.	*Viper retrovirus*

Avian (Reticuloendotheliosis) virus group:

00.061.0.02.019.	*Chick syncytial virus*
00.061.0.02.020.	*Reticuloendotheliosis virus (strain T,A)*
00.061.0.02.021.	*Trager duck spleen necrosis virus*

None reported.

Genus 00.061.0.05. *Deltaretrovirus*

Type Species 00.061.0.05.001. *Bovine leukemia virus* (BLV)

List of Species in the Genus

The ICTVdB virus code and the viruses. Official virus *species names* are in italics. Tentative virus names ( ), isolates, strains, serotypes, or subspecies are not italicized.

Virus codes, virus names, arthropod vector and host names { }, serotypes, genome sequence access

abbreviations ( ), are:

Species, their serotypes, strains and isolates

00.061.0.05.001.	*Bovine leukemia virus*	[K02120]
00.061.0.05.002.	*Primate T-lymphotropic virus 1*	
00.061.0.05.002.00.001.	*Human T-lymphotropic virus 1*	[D13784]
00.061.0.05.002.00.002.	Simian T-lymphotropic virus 1	
00.061.0.05.003.	*Primate T-lymphotropic virus 2*	
00.061.0.05.003.00.001.	*Human T-lymphotropic virus 2*	[M10060]
00.061.0.05.003.00.002.	Simian T-lymphotropic virus 2	
00.061.0.05.003.00.002.	(Formerly termed STLV-PP)	
00.061.0.05.004.	*Primate T-lymphotropic virus-3*	
00.061.0.05.004.00.002.	Simian T-lymphotropic virus 3	[Y07616]
00.061.0.05.004.00.002.	(Formerly termed STLV-L)	

Tentative Species in the Genus

None reported.

Genus 00.061.0.08. *Epsilonretrovirus*

Type Species 00.061.0.08.001. *Walleye dermal sarcoma virus* (WDSV)

List of Species in the Genus

The ICTVdB virus code and the viruses. Official virus *species names* are in italics. Tentative virus s names ( ), isolates, strains, serotypes, or subspecies are not italicized.

Virus codes, virus names, arthropod vector and host names { }, serotypes, genome sequence access abbreviations ( ), are:

Species, their serotypes, strains and isolates

00.061.0.08.001.	*Walleye dermal sarcoma virus*	[AF033822]
00.061.0.08.002.	*Walleye epidermal hyperplasia virus type 1*	[AF014792]
00.061.0.08.003.	*Walleye epidermal hyperplasia virus type 2*	[AF014793]

Tentative Species in the Genus

00.061.0.88.004.	Perch hyperplasia virus	[U26458]
00.061.0.88.005.	Snakehead retrovirus	

Genus 00.061.0.06. *Lentivirus*

Type Species 00.061.0.06.001. *Human immunodeficiency virus 1* (HIV-1)

List of Species in the Genus

The ICTVdB virus code and the viruses. Official virus *species names* are in italics. Tentative virus s names ( ), isolates, strains, serotypes, or subspecies are not italicized.

Virus codes, virus names, arthropod vector and host names { }, serotypes, genome sequence access

abbreviations ( ), are:

Species, their serotypes, strains and isolates

	Bovine lentivirus group:	
00.061.0.06.002.	*Bovine immunodeficiency virus*	[M32690]
	Equine lentivirus group:	
00.061.0.06.003.	*Equine infectious anemia virus*	[M16575]
	Feline lentivirus group:	
00.061.0.06.004.	*Feline immunodeficiency virus*	
00.061.0.06.004.03.001.	(Petuluma)	[M25381]
00.061.0.06.005.	*Feline immunodeficiency virus (Oma)*	
00.061.0.06.006.	*Puma lentivirus (PLV-14)*	
	Ovine/caprine lentivirus group:	
00.061.0.06.007.	*Caprine arthritis encephalitis virus*	[M33677]
00.061.0.06.008.	*Visna/maedi virus*	
00.061.0.06.008.04.001.	Visna/maedi virus (strain 1514)	
00.061.0.06.008.04.001.001.	VISNA strain 1514, LV1-1K51	[M60609]
00.061.0.06.008.04.001.002.	VISNA strain 1514, LV1-1K52	[M60610]
	Primate lentivirus group:	
00.061.0.06.009.	*Human immunodeficiency virus 1*	
	Several genomic clades of HIV-1 are recognised. Examples include:	
	Clade A	
00.061.0.06.009.05.001.011.	U455	[M62320]
	Clade B	
00.061.0.06.009.05.002.001.	ARV-2/SF-2	[K02007]
00.061.0.06.009.05.002.002.	BRU (LAI)	[K02013]
00.061.0.06.009.05.002.005.	HXB2	[K03455]
00.061.0.06.009.05.002.010.	RF	[M17451]
00.061.0.06.009.05.002.007.	MN	[M17449]
	Clade C	
00.061.0.06.009.05.003.013.	ETH2220	[U46016]
	Clade D	
00.061.0.06.009.05.004.008.	NDK	[M27323]
00.061.0.06.009.05.004.004.	ELI	[X04414]
	Clade F	
00.061.0.06.009.05.005.014.	93BR020	[AF005494
	Clade H	
00.061.0.06.009.05.006.015.	90CR056	[AF005496
	Clade O	
00.061.0.06.009.05.007.016.	ANT70	[L20587]
00.061.0.06.010.	*Human immunodeficiency virus 2*	
	Several genomic clades of HIV-2 are recognised. Examples include:	

Clade A

00.061.0.06.010.05.001.001.	BEN	[M30502]
00.061.0.06.010.05.001.007.	ISY	[J04498]
00.061.0.06.010.05.001.004.	ROD	[M15390]
00.061.0.06.010.05.001.006.	ST	[M31113]

Clade B

00.061.0.06.010.05.002.008.	D205	[X61240]
00.061.0.06.010.05.002.009.	EHOA	[U27200]
00.061.0.06.010.05.002.010.	UC1	[L07625]
00.061.0.06.011.	*Simian immunodeficiency virus*	
00.061.0.06.011.05.003.	African green monkey (agm) SIVs	
00.061.0.06.011.05.003.003.	African green monkey TYO	[X07805]
00.061.0.06.011.05.003.001.	African green monkey 155	[M29975]
00.061.0.06.011.05.003.002.	African green monkey 3	[M30931]
00.061.0.06.011.05.003.006.	African green monkey gr1	[M58410]
00.061.0.06.011.05.003.007.	African green monkey Sab-1	[U04005]
00.061.0.06.011.05.003.008.	African green monkey Tan-1	[U58991]
00.061.0.06.011.05.004.	chimpanzee SIV	[X52154]
00.061.0.06.011.05.005.	mandrill SIV	[M27470]
00.061.0.06.011.05.006.	red-capped mangabey SIV	[AF028607
00.061.0.06.011.05.007.	sooty mangabey SIV-H4	[X14307]
00.061.0.06.011.05.009.	*Rhesus (Maccaca mulatta)	[M19499]
00.061.0.06.011.05.008.	*pig-tailed macaque	[M32741]
00.061.0.06.011.05.010.	*stump-tailed macaque (stm)	[M83293]
00.061.0.06.011.05.011.	sykes monkey SIV	[L06042]

reported.

Genus 00.061.0.07. *Spumavirus*

Type Species 00.061.0.07.001. *Simian foamy virus 1* (SFV-1)

List of Species in the Genus

The ICTVdB virus code and the viruses. Official virus *species names* are in italics. Tentative virus s isolates, strains, serotypes, or subspecies are not italicized.

Virus codes, virus names, arthropod vector and host names { }, serotypes, genome sequence access abbreviations ( ), are:

Species, their serotypes, strains and isolates

00.061.0.07.002.	*Bovine foamy virus*	[U94514]
00.061.0.07.003.	*Chimpanzee foamy virus*	[U04327]
00.061.0.07.003.00.001.001.	*Chimpanzee foamy virus human isolate*	
00.061.0.07.003.00.001.001.	(formerly Human foamy virus)	[Y07725]
00.061.0.07.004.	*Feline foamy virus*	[Y08851]
00.061.0.07.006.	*Simian foamy virus 1*	[X54482]
00.061.0.07.007.	*Simian foamy virus 3*	[M74895]

Tentative Species in the Genus

None reported.

List of Unassigned Viruses in the Family

None reported.

Similarity with Other Taxa

None reported.

Derivation of Name

Lenti: from Latin *lentus*, "slow"
Retro: from Latin *retro*, "backwards", refers to the activity of reverse transcriptase and the transfer RNA to DNA
Spuma: from Latin *spuma*, "foam"

Contributed by

Hunter E, Casey J, Hahn B, Hayami M, Korber B, Neil J, Rethwilm A, Songio P, and Stoye J

Last Updated 30-Dec-01
Web Page by Angel Morrow
Comments to buchen@bio2.edu

3118–3129 Nucleic Acids Research, 2002, Vol. 30 No. 14

The role of template-primer in protection of reverse transcriptase from thermal inactivation

Gary F. Gerard*, R. Jason Potter, Michael D. Smith, Kim Rosenthal, Gulshan Dhariwal, Jun Lee and Deb. K. Chatterjee

Invitrogen Corporation, 7335 Executive Way, Frederick, MD 21704, USA

Received March 28, 2002; Revised and Accepted May 16, 2002

ABSTRACT

We compared the thermal stabilities of wild-type recombinant avian myeloblastosis virus (AMV) and Moloney murine leukemia virus (M-MLV) reverse transcriptase (RT) with those of mutants of the recombinant enzymes lacking RNase H activity. They differed in resistance to thermal inactivation at elevated temperatures in the presence of an RNA/DNA template-primer. RNase H-minus RTs retained the ability to efficiently synthesize cDNA at much higher temperatures. We show that the structure of the template-primer has a critical bearing on protection of RT from thermal inactivation. RT RNase H activity rapidly alters the structure of the template-primer to forms less tightly bound by RT and thus less able to protect the enzyme at elevated temperatures. We also found that when comparing wild-type or mutant AMV RT with the respective M-MLV RT, the avian enzymes retained more DNA synthetic activity at elevated temperatures than murine RTs. Enzyme, template-primer interaction again played the most significant role in producing these differences. AMV RT binds much tighter to template-primer and has a much greater tendency to remain bound during cDNA synthesis than M-MLV RT and therefore is better protected from heat inactivation.

INTRODUCTION

Reverse transcriptase (RT) is both an RNA- and DNA-directed DNA polymerase that is used extensively in recombinant DNA technology to synthesize cDNA from mRNA. Retroviral RT is encoded by the *pol* gene and is expressed as part of a polyprotein precursor that is processed into mature RT by viral-coded protease. The portion of the precursor from which RT is derived includes three structural domains arranged in the following order: NH_2-Polymerase-RNase H-Integrase-COOH (1). However, the mature structural form of RT found in the prototypic retroviruses, Moloney murine leukemia virus (M-MLV), human immunodeficiency virus (HIV) and avian sarcoma and leukosis virus (ASLV) varies both with respect to the number of structural domains and subunits present. The RT from M-MLV is a monomeric polypeptide of ~80 kDa that contains the polymerase and RNase H domains (2,3). M-MLV RT has been cloned and over expressed in *Escherichia coli* (4,5) and its RNase H-minus forms (6) have been used extensively as tools to synthesize cDNA (7,8). HIV RT is a heterodimer with a larger 66-kDa subunit having both the polymerase and RNase H domains, and a smaller 51-kDa subunit that is identical except it lacks the RNase H domain (9–11). The polymerase activity of the HIV RT heterodimer resides only in the larger subunit (12,13). Recombinant HIV RT has been studied extensively (14), but has not been used as a tool to copy mRNA because of its relatively high error rate (15). ASLV RT [includes avian myeloblastosis virus (AMV) and Rous sarcoma virus (RSV) RT] is also heterodimeric consisting of a larger 95-kDa subunit (β) and a smaller 63-kDa subunit (α) (16). ASLV RT β contains all three domains, N-terminal polymerase, RNase H and C-terminal integrase, while α contains only the polymerase and RNase H domains, the integrase domain having been deleted by proteolysis (17–19). In distinction to HIV RT, the subunits of the ASLV RT heterodimer are folded in such a way that its DNA polymerase and RNase H activities reside only on the smaller α subunit (20,21). In spite of this, the integrase domain of the larger β subunit retains enzymatic activity (19). Native, virion-derived AMV RT has also been used widely to copy mRNA (22), but until recently (23) recombinant forms of the enzyme were not available, because avian RT expressed in *E.coli* tends to be insoluble (24).

In addition to polymerase activity, RT possesses RNase H activity that degrades the RNA in an RNA/DNA hybrid. The presence of this degradative activity that is essential to the enzyme's function *in vivo* is detrimental to the synthesis of cDNA from mRNA *in vitro* (25). The RNase H domain of RT can be mutated to reduce or eliminate RNase H activity while maintaining mRNA-directed DNA polymerase activity (6,26). Removal of RT RNase H activity improves the efficiency of cDNA synthesis from mRNA catalyzed by RT (6). A second significant drawback to copying mRNA with retroviral RT is its tendency to pause during cDNA synthesis resulting in the generation of truncated products (27,28). This pausing is due in part to the secondary structure of RNA (27,29). Performing cDNA synthesis at reaction temperatures that begin to melt the secondary structure of mRNA (>55°C) helps to alleviate this problem (30). Several laboratories have demonstrated that AMV RT can be used to copy mRNA at higher temperatures

*To whom correspondence should be addressed. Tel: +1 240 379 4728; Fax: +1 240 379 4333; Email: gary.gerard@invitrogen.com

than M-MLV RT (31,32). In addition, mutations in the RNase H active site of M-MLV RT that eliminate catalytic activity appear to enhance thermal stability of RT polymerase by an unknown mechanism (33). We have compared the biochemical properties of the DNA polymerase of RNase H-plus (H+) and RNase H-minus (H–) forms of recombinant M-MLV and AMV RT in order to understand the causes of these differences in apparent thermal stability. We have found that RT interaction with the template-primer rather than intrinsic protein structural stability plays the most critical role in creating these differences. The practical implications of these differences are examined also.

MATERIALS AND METHODS

Buffers

Binding buffer was composed of 50 mM Tris–HCl pH 8.4, 75 mM KCl, 1 mM DTT and 7.5 mM $MgCl_2$. cDNA synthesis reaction buffer was identical except the DTT concentration was 10 mM. DNA denaturing buffer contained 96% (v/v) formamide, 20 mM EDTA, pH 8.0, 3 mg/ml bromphenol blue and 3 mg/ml xylene cyanol.

Materials

Native, virion-derived AMV RT was purchased from Life Sciences, Inc. SuperScript II RT (H– M-MLV RT), recombinant H+ M-MLV RT and recombinant H– and H+ RSV RT were from Invitrogen.

RNA and DNA

Synthetic RNAs (cRNAs), 1.4, 2.4, 4.4, 7.5 and 9.5 kb in length, containing a poly(A) tail at the 3′ end were obtained from Invitrogen. $(rA)_{20}$ and $(dT)_{20}$ were from Invitrogen and $(rA)_{630}$ was purchased from Miles. $p(dT)_{12-18}$, $p(dT)_{25-30}$, $(dT)_{263}$ and $(dA)_{246}$ were purchased from Amersham Pharmacia. Chloramphenicol acetyl transferase (CAT) cRNA (~900 nt), RSV cRNA (~1450 nt; derived from the *pol* gene of RSV, nt 3620–5009; 34) and MAP-4 cRNA (5.2 kb) were synthesized by T7 RNA polymerase run-off transcription from linearized plasmid DNAs (35). The sequence at the 3′ end of these cRNAs if run-off transcription was complete is 5′-$(N)_n$-$(A)_{40}$-UUAAAGUAUAUUACCA-3′. The cRNAs were selected on oligo(dT)-cellulose to ensure the presence of a poly(A) tail. The 5′ end of CAT cRNA was dephosphorylated with alkaline phosphatase. A DNA 24mer complementary to CAT cRNA that annealed between nucleotides 679 and 692 with its 5′ end 146 nt distant from the first base at the 5′ end of the CAT cRNA poly(A) tail was from Invitrogen.

CAT cDNA was synthesized from CAT cRNA in a reaction mixture (100 µl) containing cDNA synthesis reaction buffer, 500 µM each of dCTP, dATP, dTTP and [^{3}H]dGTP (36 c.p.m./pmol), 145 pmol poly(A)-tailed CAT cRNA, 145 pmol $(dT)_{20}$ and 3750 U SuperScript II. After incubation at 37°C for 15 min and 50°C for 15 min, EDTA was added to 50 mM and the reaction product was extracted with phenol/chloroform/isoamyl alcohol and ethanol precipitated. After removal of RNA by treatment with 0.3 N KOH (37°C for 16 h) and ethanol precipitation, 43 pmol of full-length single-stranded CAT cDNA was recovered. An RNA 24mer complementary to CAT cDNA that annealed between nt 679 and 692 with its 3′ end 146 nt distant from the first base at the 3′ end of the CAT cDNA oligo(dT) tail was purchased from Dharmacon Research.

Cloning and expression of AMV RT

AMV viral RNA was prepared (36) from purified AMV obtained from Life Sciences. The AMV *pol* gene cDNA was prepared from the viral RNA using the SuperScript cDNA Synthesis and Plasmid Cloning System (Invitrogen) following the manufacturer's instructions. The cDNA was cloned into plasmid pSPORT1 (Invitrogen) between the *Sal*I and *Not*I sites. Using a combination of restriction site cloning and polymerase chain reaction amplification, the gene encoding the α subunit of AMV RT was cloned into pFastBac Dual (Invitrogen) from the AMV *pol* gene in pSPORT1. The gene was cloned downstream of the p10 promoter while introducing a start codon adjacent and 5′ to nt 2796 and a stop codon adjacent and 3′ to nt 4511 (37). The gene encoding the β subunit of AMV RT was cloned into the vector containing the α gene. The β gene was cloned downstream of the polyhedrin promoter while introducing a start codon adjacent and 5′ to nt 2796 and a coding sequence for a His_6 tag followed by a 3′ stop codon downstream of nt 5369 (37).

Site-directed mutagenesis (38) was used to change the aspartic acid residue to alanine at amino acid position 450 of AMV RT (37) in both the α and β subunits. This resulted in AMV αβ RT, designated H– RT, that lacked detectable RNase H activity.

Recombinant baculovirus was formed by transforming pFastBac Dual DNA with AMV RT genes into a DH10Bac *E.coli* host (Invitrogen) bearing a low copy plasmid (bacmid) with the baculovirus genome. This resulted in transposition of the RT genes and expression control sequences into the bacmid. For expression of AMV RT, *Sf*21 cells were infected with virus at a multiplicity of infection of 2. After 72 h at 27°C, cells were harvested by centrifugation for 5 min at 2500 *g*, and cell pellets were stored at –80°C.

Purification and characterization of AMV RTs

Recombinant AMV RTs were purified from frozen infected insect cells by sequential chromatography on columns of nickle-charged chelating Sepharose (Amersham Pharmacia), AF-heparin-650 M (TosoHaas) and Mono S (Amersham Pharmacia). The enzyme was stored in 0.2 M potassium phosphate pH 7.1, 0.05% (v/v) Triton X-100, 50% (v/v) glycerol, 0.01 mM EDTA and 1 mM DTT at –80°C. As judged by SDS–PAGE, the AMV RTs were >95% homogeneous and contained equimolar amounts of α and β. They contained no detectable contaminating RNase, DNA endonuclease or DNA exonuclease. The RNA-directed DNA polymerase specific activities assayed with $(rA)_{630}op(dT)_{12-18}$ (39) of H+ and H– recombinant AMV RT were the same (57 500 U/mg) and were similar to that of native AMV RT (46 250 U/mg).

DNA polymerase assays

RT DNA polymerase unit activity was assayed with $(rA)_{630}op(dT)_{12-18}$ (39). One unit of DNA polymerase activity is the amount of RT that incorporates 1 nmol of deoxynucleoside triphosphate into acid-insoluble product at 37°C in 10 min.

cDNA synthesis from cRNA catalyzed by H– AMV RT was carried out in reaction mixtures (20 µl) containing cDNA synthesis reaction buffer, 1 mM each of dATP, dTTP, dGTP and [α-^{32}P]dCTP (250 c.p.m./pmol), 35 U RNase inhibitor, 60 nM in ends (1–3 µg) cRNA, 500 nM in ends (0.1 µg) $p(dT)_{25-30}$ and 80 nM (15 U) RT. Incubation was at 45–58°C for 30–60 min. When H+ AMV RT was used, 4 mM sodium pyrophosphate (22,40) was added and the $MgCl_2$ concentration was adjusted to 12 mM. An aliquot of the reaction mixture was precipitated with TCA to determine total yield of cDNA synthesized, and the remaining cDNA product was size fractionated on an alkaline 1.2% agarose gel (41). Reaction mixtures for H– M-MLV RT were identical except $MgCl_2$ was 3 mM, dNTPs were each 0.5 mM and RT was at 50–400 nM (25–200 U). Reaction mixtures for H+ M-MLV RT also contained 50 µg/ml actinomycin D.

To establish monovalent and divalent metal ion reaction optima, initial reaction rates were determined under conditions of limiting RT concentration during a 10 min incubation at 45°C. Reaction mixtures (20 µl) contained 3 µg (11 pmol) CAT cRNA, 0.5 µg $p(dT)_{25-30}$, 0.6 pmol RT, 50 mM Tris–HCl, pH 8.4, 10 mM DTT, 1 mM each of dATP, dTTP, dCTP and [^{3}H]dGTP (40 c.p.m./pmol), 35 U RNase inhibitor, and KCl and $MgCl_2$, varied in concentration one at a time.

To establish RT cDNA synthetic catalytic activity at elevated temperatures (>45°C), reactions were carried out in 0.5-ml tubes in a thermocycler. To ensure reactions were initiated at the desired temperature, RT was added to pre-heated reaction tubes that were not removed from the thermocycler wells during mixing.

Measurement of K_D by filter binding

A nitrocellulose filter-binding assay (42,43) was used to determine the nucleic acid binding constants (K_D) of RTs for various nucleic acids. Polynucleotides were labeled at the 5′ end with [γ-^{32}P]ATP and T4 polynucleotide kinase. Oligonucleotides were annealed to complementary polynucleotides in a buffer containing 10 mM Tris–HCl pH 7.5, and 80 mM KCl at 65°C for 5 min followed by room temperature for 15 min. The molar ratios were $(dA)_{246}o(dT)_{20}$ 1:2, CAT cRNAo$(dT)_{20}$ 1:20, CAT cRNAoDNA 24mer 1:10 and CAT cDNAoRNA 24mer 1:10. Reaction mixtures (100 µl) containing binding buffer, 0.003 or 0.03 nM labeled polynucleotide, and 0.01 to 10 nM avian RT were incubated at 23°C for 5 min. After incubation, the mixture was filtered through a nitrocellulose filter (Millipore, HA 0.45 µm) soaked in binding buffer, which was then washed with binding buffer. For M-MLV RTs, reaction mixtures contained binding buffer with the $MgCl_2$ concentration reduced to 3 mM, 0.03 nM labeled polynucleotide and 2–200 nM RT. The K_D is equal to that enzyme concentration at which 50% of the labeled polynucleotide is bound. For this method of analysis to be valid, the polynucleotide concentration in the reaction must be substantially below K_D, so that the total enzyme concentration approximates the concentration of free unbound enzyme.

Half-life determination

Mixtures (20 µl) incubated in 0.5-ml tubes in a thermocycler at 50°C contained cDNA synthesis reaction buffer and 40–80 nM RT. With murine RTs, 0.1% (v/v) Triton X-100 was also present. In some cases 200 nM CAT cRNA or 1600 nM $(rA)_{20}$ and 5400 nM $p(dT)_{12-18}$ were added. Incubation was stopped by placing tubes in ice. An aliquot (5 µl) was assayed for residual activity with $(rA)_{630}op(dT)_{12-18}$.

Reaction temperature optimum determination

RTs (1.5 U in 50-µl reaction mixture) were incubated under unit assay reaction conditions with $(rA)_{630}o(dT)_{30}$ (39) as template-primer for 5 min at temperatures ranging from 37 to 68°C. The amount of acid-insoluble DNA product synthesized at each temperature was determined.

Measurement of template breakdown catalyzed by RT RNase H

To assess the impact of RT RNase H early in a reaction on the structure of poly(A)-tailed mRNA annealed to $p(dT)_{12-18}$, CAT cRNA was labeled at the 3′ end with [α-^{32}P]ddATP (Amersham Pharmacia) and yeast poly(A) polymerase (USB) following the manufacturer's protocol. Reaction mixtures (2.5 µl) containing cDNA synthesis reaction buffer, 1.2 pmol CAT [^{32}P]cRNA (24 000 c.p.m./pmol), 27 pmol $p(dT)_{12-18}$ and 0.42 pmol avian RT were incubated for various times at 37 or 55°C. In some cases, 1 mM each of dTTP, dATP, dCTP and dGTP was present in reaction mixtures. *Escherichia coli* RNase H (1 U) was used as a positive control. Incubations were terminated by the addition of 2.5 µl of DNA denaturing buffer. After heating at 65°C for 5 min, the RNA was fractionated by electrophoresis in a 20% polyacrylamide gel containing 7 M urea.

Processivity measurement

To assess the processivity of ASLV RTs, cDNA synthesis was carried out in the presence of a heparin trap for RSV cRNA annealed to a DNA 20mer (Invitrogen) labeled at the 5′ end with ^{32}P. The DNA primer was annealed as already described to the RNA with its 3′ end 1390 nt from the 5′ end of the RNA. Reaction mixtures (2.5 µl) contained 50 mM Tris–HCl pH 8.4, 75 mM KCl, 10 mM DTT, 0.1 pmol RSV cRNA, 0.1 pmol DNA primer and 0.2 pmol ASLV RT and were pre-incubated at 45°C for 2 min. Synthesis was initiated by the addition of a solution (2.5 µl) containing Tris–HCl, KCl and DTT at the same concentrations, plus 15 mM $MgCl_2$, 2 mM each of dATP, dTTP, dGTP and dCTP and 200 mg/ml heparin. Synthesis was terminated after incubation at 45°C for 0.5 or 2 min by the addition of 5 µl of DNA denaturing buffer. A control reaction (4 µl) to test the effectiveness of the trap contained all the components listed and was incubated at 45°C for 2 min after the addition of 1 µl (0.2 pmol) of cloned H+ AMV RT to initiate the reaction. Reaction mixtures (4 µl) to assess the length of the cDNA product synthesized in the absence of trap contained all the components listed except heparin. They were incubated at 45°C for 0.5 or 2 min after addition of 1 µl (0.2 pmol) of ASLV RT. Reaction mixtures for M-MLV RT were carried out in an identical manner with the following exceptions. Incubations were at 37°C; the $MgCl_2$ and dNTP concentrations added to initiate synthesis were reduced to 6 and 1 mM, respectively; and 0.5 pmol of M-MLV RT was used. Samples were heated at 65°C for 5 min and the DNAs were fractionated by electrophoresis in a 6% polyacrylamide gel containing 7 M urea.

RESULTS

Reaction optima of recombinant AMV RT

Since recombinant forms of AMV RT have not been available for study until recently, we established reaction optima of recombinant AMV RTs and compared them with those of the native enzyme. Monovalent and divalent metal ion optima, as well as pH optima, were determined for the RNA-directed DNA polymerase activity of native AMV, cloned H+ AMV and cloned H– AMV RT utilizing CAT cRNA primed with $p(dT)_{25-30}$ (Materials and Methods). The optima for all RTs tested were identical: 75 mM KCl, 7.5 mM $MgCl_2$ and pH 8.4 in Tris–HCl buffer in the presence of 1 mM dNTPs (data not shown). Substituting Na^+ or NH_4^+ for K^+ ions did not alter the optima, nor did replacing Cl^- with CH_3COO^- salts. Because of the similarity of native and recombinant AMV RT, we focused subsequent experiments on the recombinant avian enzymes assayed under these optimal conditions.

In the process of establishing reaction optima, we also determined the reaction temperature optima of H– and H+ AMV RT. The optima of H– and H+ M-MLV RT were determined for comparison. The optima of both avian and murine RTs were much higher than expected, ranging from 43 to 54°C (Fig. 1). Strikingly, H– AMV RT still maintained near maximum relative activity even at 58°C. The optimum of each H– RT was 4°C higher than that of its respective H+ counterpart. The optimum of each avian RT was 7°C higher than that of the respective murine RT (Fig. 1). We explored the causes of these differences, and examined their practical implications.

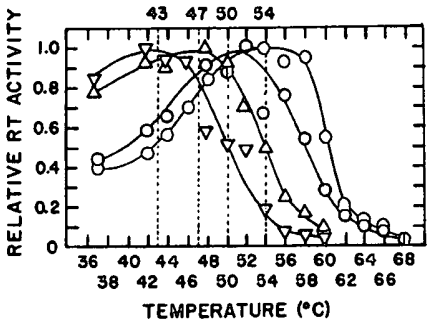


Figure 1. Reaction temperature optima of RTs. The relative RNA-directed DNA polymerase activities of H+ M-MLV RT (inverted triangle), H– M-MLV RT (triangle), H+ AMV RT (open circle) and H– AMV RT (filled circle) were measured (Materials and Methods) at the temperatures indicated. The optima (middle of the temperature range at which the relative activity was >0.9) are indicated by broken vertical lines.

Half-lives of RTs at elevated temperature

We chose to examine RT half-lives at 50°C, the maximum temperature at which AMV RT has been used to synthesize cDNA. Table 1 shows that the half-lives at 50°C of H+ and H– forms of AMV RT in the absence of template-primer were short and were similar. The same was true for H+ and H– M-MLV RT. These results indicate removal of RT RNase H activity by altering the amino acids at the active site does not change the intrinsic thermal stability of RT. Table 1 also shows that the presence of cRNAop$(dT)_{12-18}$ at 50°C increases the half-lives of most RTs substantially. The presence of such an increase supports the hypothesis that binding by RT to template-primer imparts protection from thermal inactivation. We assume that in the bound state RT is more resistant to heat inactivation than when free in solution. The increased thermal protection reflected in the increased half-life in the presence of template-primer was much greater for H– RTs. In the case of H– AMV RT, a 70-fold increase was observed. The greater increases in half-lives of H– over H+ RTs in the presence of template-primer are consistent with the optima differences in Figure 1 and could be explained by several alternatives. The first is that an increase in binding affinity for template-primer was created by the mutation introduced at the RNase H active site. This possibility seems unlikely as elimination of catalytic activity at a RT active site would be expected to leave unaltered or reduce template-primer binding. A second alternative is that RNase H activity plays a role in decreasing the thermal stability of H+ RT at elevated temperatures perhaps by altering the structure of the template-primer to a form less able to protect RT. Template-primer structure strongly influences RT binding affinity (44). These half-lives were determined in buffer containing $MgCl_2$ (Materials and Methods), so that the RNase H activity associated with H+ RT was active during the incubation and could have altered the structure of the template-primer.

Surprisingly, in contrast to the temperature optima in Figure 1, the intrinsic thermal stability of each avian RT at 50°C in the absence of template-primer was slightly less than that of its murine counterpart (Table 1). The half-lives of H– AMV and H– M-MLV RT at 37°C in the absence of template-primer were of much longer duration than at 50°C, but again the avian RT was less stable than the murine enzyme. Thus, factor(s) other than intrinsic thermal stability must be responsible for the ability of avian RTs to function at higher temperatures. There was a substantially greater increase in

Table 1. Half-lives of the DNA polymerase activity of RTs at 50°C

Enzyme	Half-life (min)[a] Minus template-primer	Plus CAT cRNAop$(dT)_{12-18}$	Plus $(rA)_{20}$op$(dT)_{12-18}$
H+ AMV	1.6 ± 0.3 (110 ± 10)[b]	15 ± 3	4 ± 0.1
H– AMV	2.1 ± 0.1 (100 ± 10)[b]	150 ± 30	15 ± 0.1
H+ M-MLV	2.8 ± 0.2 (435 ± 25)[b]	2.5 ± 0.7	ND[c]
H– M-MLV	3.2 ± 0.2 (390 ± 60)[b]	10 ± 0.2	ND[c]

[a]Mean ± standard deviation of two to four determinations.
[b]Half-lives at 37°C are shown in parentheses.
[c]ND, not determined.

Table 2. Nucleic acid dissociation constants of avian and murine RTs

Enzyme	Nucleic acid	K_D (nM)[a]
H+ AMV	$(dA)_{246}$	4 ± 0.1
	$(dA)_{246}o(dT)_{20}$	1.4 ± 0.1
	CAT cRNA	1.0 ± 0.1
H– AMV	$(dA)_{246}$	5 ± 0.8
	$(dA)_{246}o(dT)_{20}$	1.2 ± 0.3
	CAT cRNA	0.9 ± 0.1
	CAT cRNAo$(dT)_{20}$	0.1 ± 0.01
	CAT cRNAoDNA 24mer (recessed)	0.35 ± 0.01
	CAT cDNAoRNA 24mer (recessed)	0.078 ± 0.002
H+ M-MLV	CAT cRNA	29 ± 3
H– M-MLV	CAT cRNA	22 ± 7
	CAT cRNAo$(dT)_{20}$	5 ± 0.1
	CAT cRNAoDNA 24mer (recessed)	4.3 ± 0.3
	CAT cRNAoRNA 24mer (recessed)	1.0 ± 0.2

[a]Mean ± standard deviation of two or three determinations.

half-life of each avian RT over that of its murine counterpart observed in the presence of template-primer (Table 1). This differential increase is consistent with the avian RTs spending much more time bound to template-primer, perhaps by virtue of higher binding affinity.

Affinities of RTs for nucleic acids

A number of laboratories have assigned equilibrium binding constants (K_D) to HIV RT for various nucleic acids (45–50). In no case was a comparison made of H+ and H– HIV RT. H+ and H– forms of M-MLV RT were shown to have similar binding affinities for DNA/DNA template-primer (44). These K_D values were generally in the 2–10 nM range. We used a nitrocellulose filter-binding assay (42,43; Materials and Methods) to determine the equilibrium binding constants of H+ and H– RTs for various nucleic acids at room temperature (Table 2). For single-stranded $(dA)_{246}$, $(dA)_{246}$ annealed to $(dT)_{20}$ at a 1:2 M ratio, and CAT cRNA, H+ and H– AMV RT had very similar dissociation constants that were in the range reported previously for HIV RT. A comparison of binding to RNA/DNA was not done because of the potential influence of the RNase H activity of H+ RT on RNA/DNA structure. Based upon the comparisons that could be made, we conclude that alteration of the AMV RT RNase H active site by eliminating RNase H activity did not change RTs ability to bind nucleic acids. Therefore, differences in binding affinity for template-primer do not explain the higher thermal stability of H– over H+ AMV RT. There were dramatic differences observed, however, in the binding affinities of a given avian RT for DNA, DNA/DNA, RNA and RNA/DNA. Binding to single-stranded cRNA was 4-fold tighter than to single-stranded $(dA)_{246}$. Introduction of a complementary DNA primer, $(dT)_{20}$, increased the binding affinity of AMV RT for both DNA and RNA, but particularly for the 3′-poly(A) tail of CAT cRNA (9-fold increase). The order of binding affinity of AMV RT for nucleic acids was RNA/DNA >> RNA ≅ DNA/DNA > DNA. Polynucleotide length influenced binding. AMV RT binding to $(dA)_{20}$, $(rA)_{20}$, $(dA)_{20}o(dT)_{12-18}$, or $(rA)_{20}o(dT)_{12-18}$ was so weak that binding constants could not be measured accurately (K_D >250 nM).

M-MLV RTs bind with much less affinity to cRNA and cRNAo$(dT)_{20}$ than avian RTs: the K_Ds were 24–32- and 50-fold higher, respectively (Table 2). This result supports the idea that an avian RT possesses greater apparent thermal stability in the presence of template-primer than its murine RT counterpart (Table 1) because of the ability to bind much tighter to template-primer.

Model template-primer degradation studies with HIV and M-MLV RT indicate that RT prefers to bind at recessed ends of complementary oligonucleotides bound to longer oligonucleotides (51–53). When DNA with a recessed 3′ end is bound to RNA, as is the case with CAT cRNAo$(dT)_{20}$, RT binds with its polymerase active site positioned at the recessed DNA 3′ end in preparation for primer extension. When RT is exposed to RNA with recessed ends bound to DNA, RT binds with its polymerase active site positioned at the recessed RNA 5′ end, in preparation for engaging its RNase H active site in RNA cleavage (51–53). As discussed below, establishing differences in RT binding affinities for these two types of sites is important in explaining the difference in thermal stability of H– and H+ RT. To test if there is a difference in affinities, binding studies were carried out with CAT cRNA annealed to a complementary DNA 24mer and with CAT cDNA to which a RNA 24mer complement of the DNA 24mer was annealed (Materials and Methods). Both H– AMV and H– M-MLV RT bound with 4.5-fold greater affinity to the RNA/DNA hybrid with a recessed RNA 24mer than to the hybrid with a recessed DNA 24mer (Table 2). In agreement with the binding affinities established with CAT cRNAo$(dT)_{20}$, binding to these hybrids by H– AMV RT was >10-fold tighter than binding by H– M-MLV RT.

Processivities of RTs

We have already shown that H+ and H– forms of AMV RT and H+ and H– forms of M-MLV RT have similar binding affinities for nucleic acids. We also found that avian RTs bind tighter than murine RTs to nucleic acids. These affinities were determined in the absence of deoxynucleotide substrates. We wished to confirm that the equilibrium binding constants reflect the behavior of the enzymes during cDNA synthesis. Thus, the processivity of H+ and H– avian RT would be expected to be the same, and the processivities of avian RTs might be expected to be much greater than those of the murine enzymes. The processivity of a DNA polymerase can be defined as the number of nucleotides incorporated during each enzyme to template-primer binding event before the enzyme dissociates from the template-primer. The processivities of HIV, M-MLV RT and native AMV RT on cRNA annealed to a DNA primer have been reported to be relatively low, in the range of 50–100 nt (28,45,54). We compared the processivities of native and recombinant H+ and H– ASLV RTs during cDNA synthesis from an RSV cRNA template ~1400 nt long under reaction conditions established to be optimal for copying mRNA. The processivity of H+ M-MLV RT was assessed for comparison. Figure 2 (upper panel) shows that for all ASLV RTs, including H+ and H– RT, the processivity was the same and was substantially greater than 100. A heparin trap was used to restrict cDNA synthesis to a single enzyme to template-primer binding event (see Fig. 2, upper panel, lane F for control where RT was exposed to trap and template-primer simultaneously). In the presence of the trap added after RT

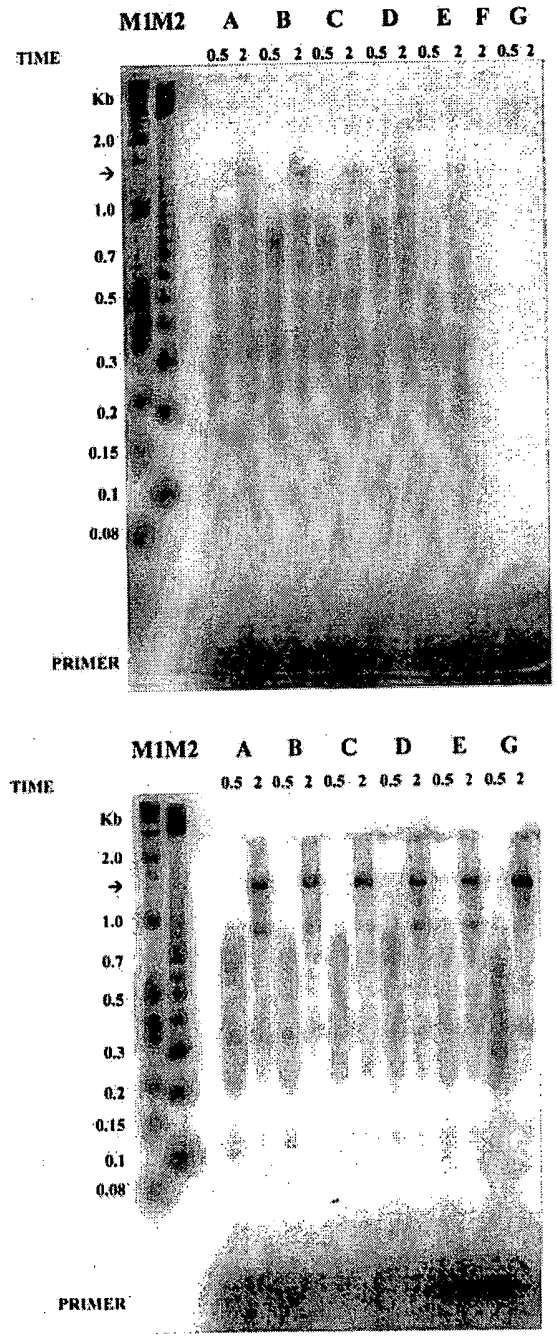


Figure 2. Processivity measurements of ASLV RTs and H+ M-MLV RT. DNA was synthesized in the presence (upper panel) and absence (lower panel) of a heparin trap from RSV cRNA annealed to a 5′ ^{32}P-labeled DNA 20mer (Materials and Methods). Reaction mixtures were incubated for 0.5 and 2 min and contained cloned H+ AMV RT (A), native AMV RT (B), H– AMV RT (C), cloned H+ RSV RT (D), H– RSV RT (E) or H+ M-MLV RT (G). (F, upper panel) The total inhibition of cDNA synthesis by cloned H+ AMV RT initiated in the presence of both template-primer and heparin, demonstrating the effectiveness of the heparin trap. ^{32}P-labeled markers were 1 kb DNA ladder (lane M1) and 100 bp DNA ladder (lane M2). The arrow indicates full-length product.

was permitted to bind to template-primer, avian RTs synthesized cDNA product 90–900 nt long in 30 s and 90 nt to full length (~1400 nt) in 2 min at 45°C (Fig. 2, upper panel, A–E). Approximately one-half of the cDNA product made by each RT was >400 nt long. At 45°C both H+ and H– avian RT catalyze cDNA synthesis with an average processivity >400 nt and a chain growth rate of 20–30 nt/s. Processivity measured with a longer cRNA template showed the avian RTs do not synthesize cDNA products much beyond 1400 nt long during a single cycle of synthesis (data not shown). Nucleotide concentration influenced processivity. These experiments were carried out in the presence of 1 mM dNTPs. At 0.5 mM dNTPs, there was no change in the processivity (data not shown). At 0.2, 0.1 and 0.02 mM dNTPs, however, the maximum length of cDNA product synthesized by cloned AMV RT from the RSV cRNA 1400-nt template in the presence of trap dropped from ~1400 to ~1000, ~500 and ~200 nt, respectively (data not shown). If multiple initiation events were permitted to occur in the absence of trap (Fig. 2, lower panel), most cDNA products were full-length in 2 min. Exceptions were defined cDNA products ~350 and ~900 nt long that were observed in the presence of trap and persisted in the absence of trap. Apparently the avian RT terminates processive synthesis at a number of sites, but is also capable of continued synthesis uninterrupted to the end of the template. At a few of these pause sites, such as at template positions corresponding to products ~350 and ~900 nt long, the enzyme has greater difficulty reinitiating DNA synthesis. H+ M-MLV RT synthesized little discernible cDNA product longer than primer in the presence of trap (Fig. 2, upper panel, G). In the absence of trap, H+ M-MLV RT synthesized mostly full-length product (Fig. 2, lower panel, G), indicating that the reaction conditions used supported efficient cDNA synthesis. Analysis by denaturing polyacrylamide gel electrophoresis of the length of cDNA product synthesized by H+ or H– M-MLV RT shows their processivities are 20–40 nt (data not shown). We conclude that differences in template-primer binding affinity and processivity do not explain the higher apparent thermal stability of H– over H+ RT, but do play a critical role in making avian RT more thermal stable than murine RT in the presence of template-primer.

Changes in template-primer structure produced by RT RNase H

To explain the higher apparent thermal stability of H– over H+ RT, we are left with the possibility that RT RNase H alters template-primer structure in a fashion that reduces RT binding affinity and thus protection from heat inactivation. The results in Table 3 show that exposure of CAT cRNA•$(dT)_{12-18}$ to the RNase H activity of H+ AMV RT drastically diminishes the ability of this template-primer to protect H– AMV RT from thermal inactivation. The half-life of H– AMV RT at 50°C in the presence of CAT cRNA•$(dT)_{12-18}$ that received no pre-treatment with RT was 150 min. Pre-exposure to H– AMV RT had little effect on the ability of CAT cRNA•$(dT)_{12-18}$ to protect H– RT, reducing the half-life only slightly to 135 min. In contrast, pre-treatment of template-primer with H+ AMV RT reduced the half-life of H– RT 9-fold to 19 min.

By labeling a CAT cRNA template at its 3′ end, we were able to assess the impact the RNase H activity of AMV RT had

Table 3. Half-life of the DNA polymerase activity of H– AMV RT at 50°C in the presence of pre-treated template-primer

Pre-treatment of template-primer	Half-life (min)[a]
None[b]	150 ± 30
RNase H– RT[c]	135 ± 10
RNase H+ RT[d]	19 ± 1

[a]Mean ± standard deviation of two or three determinations.
[b]The half-life of H– AMV RT (1.6 pmol) was determined in the presence of CAT cRNA•p(dT)$_{12-18}$ (4/108 pmol) that received no pre-treatment.
[c]Pre-treatment with 1.6 pmol of H– AMV RT in the presence of Mg^{2+} at 55°C for 5 min and 65°C for 10 min.
[d]Pre-treatment with H+ AMV RT in the presence of Mg^{2+} at 55°C for 5 min and 65°C for 10 min (incubation at 65°C inactivated RT).

on the template-primer structure that apparently is protecting the RT. The CAT cRNA used in these studies contained a 3′-poly(A) tail of ~50 nt (Materials and Methods). If the poly(A) tail is removed by RT RNase H one would expect to see a distribution of labeled cleavage products 50 nt or less in length. A high incubation temperature (55°C) was used to assess whether p(dT)$_{12-18}$•cRNA is indeed a substrate for RT RNase H even at temperatures near the upper limit of RT use. Polyacrylamide gel electrophoresis was used to fractionate any breakdown products generated from 3′ ^{32}P-labeled CAT cRNA annealed to p(dT)$_{12-18}$ after a brief exposure to AMV RT (Fig. 3). As might be expected, incubation with H– AMV RT did not discernibly change the structure of the cRNA (Fig. 3, compare lanes H and I with lanes B and C). In contrast, incubation with H+ AMV RT for even a brief period (15 s) resulted in rapid cleavage of the cRNA poly(A) tail with formation of oligomers 5–50 nt long (Fig. 3, lane D). The product size distribution was the same at 2 min (Fig. 3, lane E). At 15 s and 2 min, only 19 and 8%, respectively, of the radioactivity remained in undigested template. Incubation for longer periods with a greater amount of H+ RT reduced the length distribution to a limit digest of 2–30 nt (Fig. 3, lane G). Similar results were obtained with 8 pmol of H+ M-MLV RT at 37°C (data not shown). So the action of RT RNase H on cRNA•(dT)$_{12-18}$ at the beginning of a cDNA synthesis reaction results in deadenylation of the cRNA with the formation of single-stranded cRNA and (rA)$_n$•(dT)$_{12-18}$ hybrids with n varying from 5 to 50 nt. Making the assumption that (rA)$_{20}$•p(dT)$_{12-18}$ mimics the structure of many of the cleavage products removed from the 3′ end of CAT cRNA•p(dT)$_{12-18}$ by RT RNase H, the results in Table 1 (last column on the right) indicate that (rA)$_n$•p(dT)$_{12-18}$ protects RT poorly from heat inactivation. Also, single-stranded RNA protects RT poorly from thermal inactivation based upon the results in Table 3. Consistent with reduced thermal protection of RT by (rA)$_n$•(dT)$_{12-18}$ or single-stranded RNA, RT binds with much lower affinity to these nucleic acids than to cRNA•DNA (Table 2 and text). We conclude that in the presence of RNase H the structure of the template-primer is changed quite rapidly at the beginning of a reaction to a deadenylated form that does not protect RT well from thermal inactivation.

The results in Figure 3 were obtained in the absence of dNTPs, when RNase H activity is independent of polymerase

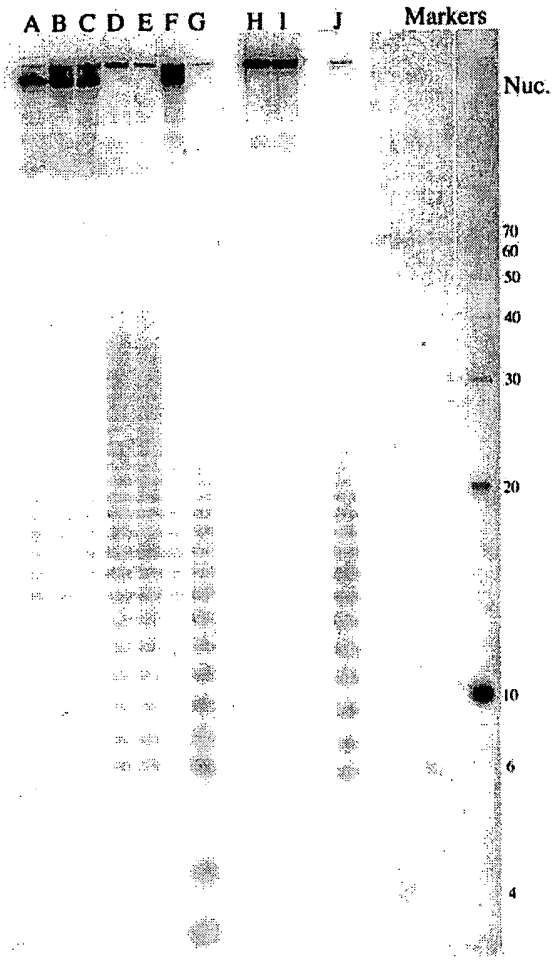


Figure 3. Deadenylation of CAT cRNA•p(dT)$_{12-18}$ by AMV RT. CAT cRNA ^{32}P-labeled at the 3′ end and annealed to p(dT)$_{12-18}$ was incubated with H+ and H– AMV RT for various times at 37 or 55°C and then fractionated on a denaturing 20% polyacrylamide gel (Materials and Methods). Template-primer that was not incubated is shown in lane A. Samples were incubated at 4°C for 2 min without RT (lane B); 55°C for 2 min without RT (lane C); 55°C for 15 s (lane D) and 2 min (lane E) with 0.42 pmol H+ AMV RT; 37°C for 5 min without RT (lane F); 37°C for 5 min with 4.2 pmol H+ AMV RT (lane G); 55°C for 15 s (lane H) and 2 min (lane I) with 0.42 pmol H– AMV RT; and 37°C for 5 min with 1 U *E.coli* RNase H (lane J). ^{32}P-labeled markers were 10 bp ladder, (dT)$_4$ and (dT)$_6$.

activity. In the presence of dNTPs, the results were similar. Oligomers 5–50 nt long were again produced and very little intact RNA persisted after 15 s. In addition, some larger breakdown products 70–200 nt long were observed at 15 s that were reduced to smaller species (5–50 nt) in 2 min (data not shown).

Practical implications

A suitable method for judging the useful upper temperature limit of the DNA polymerase activity of RT is an assessment of the effect of increasing reaction temperature on the amounts of full-length cDNA products synthesized by RT from an equimolar mixture of cRNAs of various lengths. The labeled,

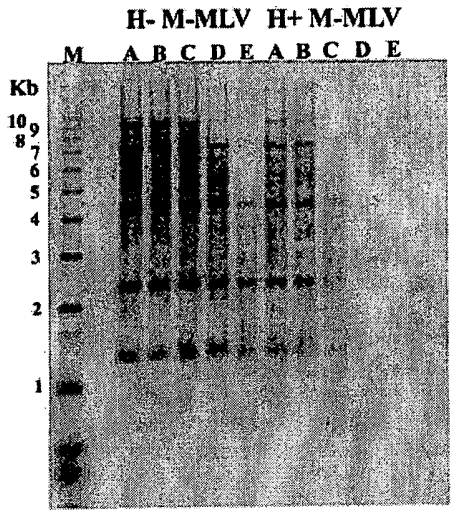


Figure 4. The effect of temperature on full-length cDNA synthesis from equimolar amounts of cRNAs. cDNAs synthesized by H– and H+ M-MLV RT (400 nM) at 42°C (lane A), 45°C (lane B), 48°C (lane C), 50°C (lane D) and 52°C (lane E) from equimolar amounts (12 nM each) of cRNAs of 1.4 (0.1 μg), 2.4 (0.17 μg), 4.4 (0.31 μg), 7.5 (0.54 μg) and 9.5 kb (0.67 μg) were analyzed by alkaline agarose gel electrophoresis (Materials and Methods). Incubation time was 30 min. ^{32}P-labeled DNA (1 kb ladder) was run as a marker (lane M).

full-length cDNA products can be separated and quantified on an alkaline agarose gel. As the reaction temperature is increased, full-length cDNA products disappear starting with those derived from longer cRNAs until a temperature is reached where no discernible full-length product of any length is synthesized. The results of such a gel analysis carried out with H– and H+ M-MLV RT at 42–52°C are shown in Figure 4. H– M-MLV RT continues to synthesize cDNAs from cRNAs 1.4–9.5 kb in length up to 50°C, while H+ M-MLV RT does so only up to 45°C. The results of this analysis and a similar analysis of full-length cDNA products synthesized by AMV RTs at temperatures ranging from 45 to 58°C are summarized in Table 4. At temperatures up to 50°C, H+ AMV RT and H– AMV RT synthesized full-length cDNAs from cRNAs 1.4–9.5 kb in length. At temperatures >50°C, only H– AMV RT continued to make full-length products of all lengths.

The amounts of enzymes used to synthesize cDNA in Figure 4 and Table 4 were established empirically to be those required to synthesize the maximum amount of long cDNA (>5 kb) at temperatures (37°C) and incubation times (60 min) normally used to prepare cDNA (data not shown). At 37°C, both avian and murine RTs are stable for extended periods (Table 1) so that enzyme stability should not be a factor in determining the amount of enzyme required. In spite of this, AMV RT was required only in near stoichiometric amounts (80 nM RT with 60 nM total template), while M-MLV RTs were required in much higher amounts (400 nM). Table 5 shows that when either H– or H+ M-MLV RT is used at 37°C during a 60-min incubation in an amount equivalent to a 5.2-kb cRNA template (50 nM), very little full-length cDNA is synthesized. Changing to a long incubation time partially alleviates this diminishment in full-length cDNA synthesis, as demonstrated by the increase in full-length products synthesized by 50 nM H– M-MLV RT during a 240-min incubation (Table 5). So an RT possessing relatively low processivity and nucleic acid binding affinity must be used in excess

Table 4. Functional cDNA synthetic activity of RTs at elevated temperatures

Enzyme	Temperature (°C)	Amount of full-length product (ng)[a]				
		1.4 kb	2.4 kb	4.4 kb	7.5 kb	9.5 kb
H+ AMV[c]	45	17	26	36	38	31
	50	12	17	21	15	10
	55	10	10	4	<2[b]	<2
	58	<2	<2	<2	<2	<2
H– AMV[c]	45	26	39	55	63	44
	50	21	32	45	47	30
	55	18	28	36	30	12
	58	8	13	14	9	3
H+ M-MLV[c]	42	26	32	29	23	7
	45	24	28	24	13	4
	48	16	17	5	<2	<2
	50	<2	<2	<2	<2	<2
H– M-MLV[c]	42	30	40	43	51	30
	45	27	34	38	49	33
	48	37	47	50	59	24
	50	30	39	38	25	4
	52	25	26	7	<2	<2

[a]cDNA synthesis reaction mixtures (see Materials and Methods) contained equimolar amounts (12 nM each; 60 nM total) of the following cRNAs: 1.4 (0.1 μg), 2.4 (0.17 μg), 4.4 (0.31 μg), 7.5 (0.54 μg) and 9.5 kb (0.67 μg). The amounts of full-length product were established by cutting the region in a dried 1.2% alkaline agarose gel corresponding to the size of each full-length band and counting it in a scintillation counter. The results of a single experiment are shown. Similar results were obtained in at least two other separate experiments.

[b]Only amounts of full-length product >2 ng could be seen as discernible bands on the gel autoradiograph at the exposure times used.

[c]Reactions contained 80 nM AMV RTs or 400 nM M-MLV RTs.

Table 5. cDNA synthesis by RTs at different enzyme concentrations at 37°C

Enzyme	Enzyme concentration (nM)	Amount of full length product (ng)[a]
H+ AMV	25	37 ± 2
	50	71 ± 14
	400	100 ± 30
H– AMV	25	44 ± 16
	50	73 ± 2
	400	110 ± 21
H+ M-MLV	25	<2
	50	<2
	400	70 ± 13
H– M-MLV	25	<2
	50	15 ± 7 (56 ± 4)
	400	130 ± 16

[a]cDNA synthesis reaction mixtures (see Materials and Methods) contained 50 nM (1 μg) 5.2 kb cRNA. Incubation was at 37°C for 60 min, with the exception of the value in parentheses that was determined after a 240-min incubation. The amounts of full-length product were established by cutting the region in a dried 1.2% alkaline agarose gel corresponding to the size of full-length product and counting in a scintillation counter. The mean ± standard deviation of two to four determinations is shown.

amounts or for long incubation periods to synthesized full-length cDNA product from long mRNA. As the incubation temperature of a reaction is increased >37°C, the requirement for excess amounts of M-MLV RT becomes even greater.

DISCUSSION

Stabilization of enzymes to thermal inactivation by binding to substrate is a well-documented phenomenon (55) that is thought to take place because of substrate-induced conformational changes in a monomeric enzyme and/or enhanced subunit interaction in a multimeric enzyme (56). Mutations that enhance substrate binding therefore often increase thermal stability. The mutations introduced into AMV RT (D450A) and M-MLV RT (D524G, E562Q and D583N; unpublished data) that eliminated RNase H catalytic activity probably did so by eliminating binding of the divalent metal ion(s) required for activity (57–59). However, we have demonstrated that these mutations do not affect RT template-primer binding affinity or processivity. The apparent thermal stabilization of RT introduced by altering the RNase H active site is a new and different phenomenon caused by virtue of maintaining RTs substrate in a particular structural state.

The thermal protection data in Table 1 demonstrates poly(A)-tailed mRNA mixed with an excess of $p(dT)_{12-18}$ protects H– AMV RT effectively in the absence of dNTPs at 50°C. The enzyme apparently has a preference for binding tightly to oligo(dT) annealed to the poly(A) tail of mRNA (Table 2). AMV H– RT efficiently initiates cDNA synthesis from mRNAop$(dT)_{10}$ or mRNAop$(dT)_{20}$ at temperatures in the 50–55°C range, above the predicted T_ms of the poly(rA)/oligo(dT) hybrids (G.Gerard, unpublished data). This suggests that bound AMV RT helps stabilize the duplex structure of oligo(dT)opoly(A)mRNA. In a typical cDNA synthesis reaction containing 15 U AMV RT (1.6 pmol), 1 μg of mRNA (1.6 pmol of RNA ends if the average mRNA is 2 kb), and an excess of $p(dT)_{12-18}$ [10–20 pmol of $p(dT)_{12-18}$ primer ends bound per mRNA ~$(A)_{200}$ tail], there is an excess of a single type of RT tight binding site available. This site is composed of DNA with a recessed 3′ end bound to a longer RNA. At this type of binding site the RT polymerase domain is positioned at the DNA recessed 3′-terminus in preparation for initiation of cDNA synthesis (51–53). In the absence of RNase H activity, this type of binding site remains available throughout the course of the reaction to protect RT from thermal inactivation. The mRNA poly(A) tail remains intact and is available for unused $p(dT)_{12-18}$ primers to remain bound, providing DNA recessed 3′ end binding sites. The 3′-OH end of the growing cDNA chain also persists as a recessed 3′ end binding site. This is not the case in the presence of RT RNase H activity. During initiation of cDNA synthesis from mRNA there is opportunity for RT RNase H to remove the mRNA poly(A) tail in a polymerase-dependent and -independent manner (28,44,60–63). Initiation of cDNA synthesis and RT RNase H-catalyzed cleavage of the mRNA poly(A) tail appear to be almost instantaneous with initiation of cDNA synthesis apparently occurring first (Fig. 3; G.Gerard, unpublished data). The poly(A) tail of mRNAooligo(dT) is quickly converted by RT RNase H into oligo(rA) fragments (Fig. 3), many of which are too short to remain bound to oligo(dT). Oligo(dT), oligo(rA) and oligo(dT)ooligo(rA) hybrids are bound with much lower affinity by RT and protect the enzyme poorly from heat inactivation during the course of the reaction (Table 1). In addition the RNase H activity of H+ RT catalyzes polymerase-dependent cleavage of RNA as cDNA synthesis proceeds, generating RNA oligonucleotides 6–20 nt long (28,44,60–63). The RNA oligonucleotides that remain bound to cDNA product quickly become the predominant potential RT binding sites in the reaction mixture. The binding affinity studies in Table 2 show that this type of binding site is actually preferred over the type of site present at the 3′ end of the growing cDNA chain. RT that disengages from cDNA synthesis because of pausing might be expected to bind preferentially at these RNA recessed ends. This would have several detrimental effects. First, enzyme engaged in productive cDNA synthesis would be diminished. Secondly, at these sites RT binds in such a way that the polymerase domain is positioned at the 5′ end of the RNA hybridized to the longer cDNA (51–53), preparing the enzyme for polymerase-independent cleavage of the RNA. Such cleavage would result in the generation of small RNA oligonucleotides that do not remain bound to DNA and in the loss of the hybrid structure bound by RT, diminishing the ability of the template-primer to protect RT from thermal inactivation.

The situation with M-MLV RTs is somewhat different. The enhancement in template-primer thermal protection obtained by eliminating the RNase H activity of M-MLV RT was not nearly as great as obtained with AMV RT (Table 1). Optimal reaction temperatures were lower for the murine RTs (Fig. 1 and Table 4), and the enhancement in the upper temperature limit of effective use was not as great for murine RTs (Table 4). H– M-MLV RT could be used at 50°C, as compared with 45°C for H+ M-MLV, while H– AMV RT could be used at 58°C, as compared with 50°C for H+ AMV RT. These differences were not due to intrinsic thermal stability as the murine RTs actually had slightly greater half-lives than the avian enzymes at 50°C in the absence of nucleic acids (Table 1). The dominant factor in causing these differences is

the weaker interaction between murine RT and template-primer. Avian RTs bind with much higher affinity and their polymerases have greater processivity than murine RTs. The K_D of AMV H– RT is 50-fold lower than that of M-MLV H– RT to cRNAo(dT)$_{20}$ (Table 2), and processivities of the avian enzymes are at least 10-fold greater than murine RTs at high dNTP concentration. Therefore, during cDNA synthesis murine RT spends much more time than avian RT in an unbound state not protected by template-primer from thermal inactivation. Also consistent with the data in Table 5, a less processive RT that spends more time cycling between a bound catalytically active state and a non-productive unbound state would be expected to be more dependent on the presence of excess enzyme to synthesize full-length copies of longer mRNAs.

Heteropolymeric RNA-directed DNA polymerase processivity values reported for HIV, M-MLV and AMV RT are relatively low (average of 50–100 nt) (28,45,54). We found the processivity of the polymerase activity of all ASLV RTs to be much higher, averaging >400 nt with an upper limit of ~1400 nt (Fig. 2). This discrepancy can be explained at least in part by the fact that prior processivity measurements were made at dNTP concentrations of between 25 and 150 μM. Nucleotide concentrations in this range match or exceed by several fold K_{mdNTP} values reported for avian and murine RTs (46,64,65). The processivity assays described here were performed at 0.5–1 mM dNTPs, a concentration range reported to be more optimal for synthesis of cDNA from long mRNA (40,66,67). Decreasing the dNTP concentration reduced dramatically the processivity of ASLV αβ RT polymerase from an upper limit of ~1400 nt at 1 mM dNTPs to ~200 nt at 20 μM dNTPs. In contrast, dNTP concentration appears not to influence the processivity of murine RTs to the same extent. At 25 μM dNTPs, the value reported for M-MLV RT polymerase was 69 nt (28), and we found at 500 μM dNTPs the processivity was ~30 nt for H– and H+ M-MLV RT. The DNA polymerase of avian αβ RT is therefore much more processive than murine RT polymerase at high dNTP concentrations. While at dNTP concentrations approaching the K_{mdNTP}, avian RT processivity begins to approach that of the murine enzyme. The β subunit of ASLV αβ RT possesses an additional nucleic acid binding domain, integrase, not present in HIV or M-MLV RT (19,68). The presence of this domain increases the affinity of ASLV αβ RT for nucleic acids and increases the processivity of its polymerase relative to the α form (23,64) that is missing the domain. We speculate that high dNTP concentrations increase the binding affinity of avian αβ RT for template-primer and thus increase processivity. Perhaps dNTPs at high concentrations are bound by the ASLV RT integrase domain resulting in an increase in binding affinity of the enzyme for template-primer. High concentrations (2.5 mM) of nucleoside triphosphate have been reported to stimulate the catalytic activity of AMV RT integrase (69).

Maintenance of polymerase activity at elevated temperatures has practical implications for the use of RT in copying mRNA. The secondary structure of mRNA is reduced as temperatures are increased, reducing the chances that RT will terminate cDNA synthesis at secondary structural features in the RNA (29,63). The ability to perform cDNA synthesis at elevated temperatures also improves the specificity with which the reaction is primed (70).

We have shown in this study that tight binding by RT to DNA recessed 3′ ends in template-primer protects RT from thermal inactivation. One approach demonstrated here to achieving higher incubation temperatures with RT is to maintain this type of binding site in the template-primer during the course of the reaction. An alternative approach to increasing retroviral RT thermal stability is to make amino acid changes in RT that increase binding to template-primer and/or increase the enzyme's intrinsic thermal stability.

REFERENCES

1. Prasad,V.R. (1993) Genetic analysis of retroviral reverse transcriptase structure and function. In Skalka,A.M. and Goff,S.P. (eds), *Reverse Transcriptases*. CSHL Press, Plainview, NY, pp. 135–162.
2. Moelling,K. (1974) Characterization of reverse transcriptase and RNase H from Friend-murine leukemia virus. *Virology*, **62**, 46–59.
3. Gerard,G.F. and Grandgenett,D.P. (1975) Purification and characterization of the DNA polymerase and RNase H activities in Moloney murine sarcoma-leukemia virus. *J. Virol.*, **15**, 785–797.
4. Kotewicz,M.L., D'Alessio,J.M., Driftmier,K.M., Blodgett,K.P. and Gerard,G.F. (1985) Cloning and overexpression of Moloney murine leukemia virus reverse transcriptase in *Escherichia coli*. *Gene*, **35**, 249–258.
5. Roth,M.J., Tanese,N. and Goff,S.P. (1985) Purification and characterization of murine retroviral reverse transcriptase expressed in *Escherichia coli*. *J. Biol. Chem.*, **260**, 9326–9335.
6. Kotewicz,M.L., Sampson,C.M., D'Alessio,J.M. and Gerard,G.F. (1988) Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H. *Nucleic Acids Res.*, **16**, 265–277.
7. Gerard,G.F. and D'Alessio,J.M. (1993) Reverse transcriptase. The use of cloned Moloney murine leukemia virus reverse transcriptase to synthesize DNA from RNA. In Burrell,M.M. (ed.), *Methods in Molecular Biology, Vol. 16: Enzymes of Molecular Biology*. Humana Press, Totowa, NJ, pp. 73–93.
8. Gerard,G.F., Fox,D.K., Nathan,M. and D'Alessio,J.M. (1997) Reverse transcriptase. The use of cloned Moloney murine leukemia virus reverse transcriptase to synthesize DNA from RNA. *Mol. Biotechnol.*, **8**, 61–77.
9. DiMarzo Veronese,F., Copeland,T.D., Devico,A.L., Rahman,R., Oroszlan,S., Gallo,R.C. and Sarngadharan,M.G. (1986) Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. *Science*, **231**, 1289–1291.
10. Hansen,J., Schulze,T. and Moelling,K. (1987) RNase H activity associated with bacterially expressed reverse transcriptase of human T-cell lymphotropic virus III/lymphadenopathy-associated virus. *J. Biol. Chem.*, **262**, 12393–12396.
11. Hansen,J., Schulze,T., Mellert,W. and Moelling,K. (1988) Identification and characterization of HIV-specific RNase H by monoclonal antibody. *EMBO J.*, **7**, 239–243.
12. LeGrice,S.F.J., Naas,T., Wohlgensinger,B. and Schatz,O. (1991) Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1 reverse transcriptase. *EMBO J.*, **10**, 3905–3911.
13. Hostomsky,Z., Hostomska,Z., Fu,T. and Taylor,J. (1992) Reverse transcriptase of human immunodeficiency virus type 1: functionality of subunits of the heterodimer in DNA synthesis. *J. Virol.*, **66**, 3179–3182.
14. LeGrice,S.F.J. (1993) Human immunodeficiency virus reverse transcriptase. In Skalka,A.M. and Goff,S.P. (eds), *Reverse Transcriptases*. CSHL Press, Plainview, NY, pp. 163–191.
15. Bebenek,K. and Kunkel,T.A. (1993) The fidelity of retroviral reverse transcriptases. In Skalka,A.M. and Goff,S.P. (eds), *Reverse Transcriptases*. CSHL Press, Plainview, NY, pp. 85–102.
16. Grandgenett,D.P., Gerard,G.F. and Green,M. (1973) A single subunit from avian myeloblastosis virus with both RNA-directed DNA polymerase and ribonuclease H activity. *Proc. Natl Acad. Sci. USA*, **70**, 230–234.
17. Rho,H.M., Grandgenett,D.P. and Green,M. (1975) Sequence relatedness between the subunits of avian myeloblastosis virus reverse transcriptase. *J. Biol. Chem.*, **250**, 5278–5280.
18. Schiff,R.D. and Grandgenett,D.P. (1978) Virus-coded origin of a 32,000-dalton protein from avian retrovirus cores: structural relatedness of p32

and the beta polypeptide of the avian retrovirus DNA polymerase. *J. Virol.*, **28**, 279–291.
19. Golomb,M. and Grandgenett,D.P. (1979) Endonuclease activity of purified RNA-directed DNA polymerase from avian myeloblastosis virus. *J. Biol. Chem.*, **254**, 1606–1613.
20. Werner,S. and Wohrl,B.M. (2000) Asymmetric subunit organization of heterodimeric Rous sarcoma virus reverse transcriptase αβ: localization of the polymerase and RNase H active sites in the α subunit. *J. Virol.*, **74**, 3245–3252.
21. Werner,S. and Wohrl,B.M. (2000) Homodimeric reverse transcriptase from Rous sarcoma virus mutated within the polymerase and RNase H active site of one subunit are active. *Eur. J. Biochem.*, **267**, 4740–4744.
22. Kimmel,A.R. and Berger,S.L. (1987) Preparation of cDNA and the generation of cDNA libraries: overview. *Methods Enzymol.*, **152**, 307–316.
23. Werner,S. and Wohrl,B.M. (1999) Soluble Rous sarcoma virus reverse transcriptases alpha, alphabeta and beta purified from insect cells are processive DNA polymerases that lack an RNase H 3′ to 5′ directed processing activity. *J. Biol. Chem.*, **274**, 26329–26336.
24. Soltis,D.A. and Skalka,A.M. (1988) The alpha and beta chains of avian retrovirus reverse transcriptase independently expressed in *Escherichia coli*: characterization of enzymatic activities. *Proc. Natl Acad. Sci. USA*, **85**, 3372–3376.
25. Berger,S.L., Wallace,D.M., Puskas,R.S. and Eschenfeldt,W.H. (1983) Reverse transcriptase and its associated ribonuclease H: interplay of two enzyme activities controls the yield of single-stranded complementary deoxyribonucleic acid. *Biochemistry*, **22**, 2365–2372.
26. DeStefano,J.J., Wu,W., Seehra,J., McCoy,J., Laston,D., Alborne,E., Fay,P.J. and Bambara,R.A. (1994) Characterization of an RNase H deficient mutant of human immunodeficiency viurs-1 reverse transcriptase having an aspartate to asparagine change at position 498. *Biochim. Biophys. Acta*, **1219**, 380–388.
27. Harrison,G.P., Mayo,M.S., Hunter,E. and Lever,A.M.L. (1998) Pausing of reverse transcriptase on retroviral RNA templates is influenced by secondary structures both 5′ and 3′ of the catalytic site. *Nucleic Acids Res.*, **26**, 3433–3442.
28. DeStefano,J.J. Buiser,R.G., Mallaber,L.M., Myers,T.W., Bambara,R.A. and Fay,P.J. (1991) Polymerization and RNase H activities of reverse transcriptases from avian myeloblastosis, human immunodeficiency and Moloney murine leukemia viruses are functionally uncoupled. *J. Biol. Chem.*, **266**, 7423–7431.
29. Wu,W., Henderson,L.E., Copeland,T.D., Gorelick,R.J., Bosche,W.J., Rein,A. and Levin,J.G. (1996) Human immunodeficiency virus type 1 nucleocapsid protein reduces reverse trancriptase pausing at a secondary structure near the murine leukemia virus polypurine tract. *J. Virol.*, **70**, 7132–7142.
30. Myers,T.W. and Gelfand,D.H. (1991) Reverse transcription and DNA amplification by *Thermus thermophilus* DNA polymerase. *Biochemistry*, **30**, 7661–7666.
31. Tosh,C., Hemadri,D., Sanyal,A., Pattnaik,B. and Venkataramanan,R. (1997) One-tube and one-buffer system of RT–PCR amplification of 1D gene of foot-and-mouth disease virus field isolates. *Acta Virol.*, **41**, 153–155.
32. Freeman,W.M., Vrana,S.L. and Vrana,K.E. (1996) Use of elevated reverse transcriptase reaction temperatures in RT–PCR. *Biotechniques*, **20**, 782–783.
33. Gerard,G.F., Schmidt,B.J., Kotewicz,M.L. and Campbell,J.H. (1992) cDNA synthesis by Moloney murine leukemia virus RNase H-minus reverse transcriptase possessing full DNA polymerase activity. *Focus*, **14**, 91–93.
34. Schwartz,D.E., Tizard,R. and Gilbert,W. (1983) Nucleotide sequence of Rous sarcoma virus. *Cell*, **32**, 853–869.
35. D'Alessio,J.M. and Gerard,G.F. (1988) Second-strand cDNA synthesis with *E.coli* DNA polymerase I and RNase H: the fate of information at the mRNA 5′ terminus and the effect of *E.coli* DNA ligase. *Nucleic Acids Res.*, **16**, 1999–2014.
36. Strauss,E.M., Simon,M.C. and Kimball,P.C. (1980) Efficient production of mammalian RNA tumor viruses in serum-free culture medium allows rapid RNA subunit purification. *J. Virol. Methods*, **1**, 213–221.
37. Joliot,V., Boroughs,K., Lasserre,F., Crochet,J., Dambrine,G., Smith,R.E. and Perbal,B. (1993) Pathogenic potential of myeloblastosis-associated virus: implications of env proteins for osteopertrosis induction. *Virology*, **195**, 812–819.
38. Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.*, **154**, 367–382.
39. Houts,G.E., Miyagi,M., Ellis,C., Beard,D. and Beard,J.W. (1979) Reverse transcriptase from avian myeloblastosis virus. *J. Virol.*, **29**, 517–522.
40. Myers,J.C., Spiegelman,S. and Kacian,D.L. (1977) Synthesis of full-length DNA copies of avian myeloblastosis virus RNA in high yields. *Proc. Natl Acad. Sci. USA*, **74**, 2840–2843.
41. McDonell,M.W., Simon,M.N. and Studier,F.W. (1977) Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *J. Mol. Biol.*, **110**, 119–146.
42. Bailey,J.M. (1979) Interpretation of nitrocellulose filter assays of protein-nucleic acid binding. *Anal. Biochem.*, **93**, 204–206.
43. Strauss,H.S., Boston,R.S., Record,M.T.,Jr and Burgess,R.R. (1981) Variables affecting the selectivity and efficiency of retention of DNA fragments by *E.coli* RNA polymerase in the nitrocellulose-filter-binding assay. *Gene*, **13**, 75–87.
44. Schultz,S.J., Zhang,M., Kelleher,C.D. and Champoux,J.J. (1999) Polypurine tract primer generation and utilization by Moloney murine leukemia virus reverse transcriptase. *J. Biol. Chem.*, **274**, 34547–34555.
45. Huber,H.E., McCoy,J.M., Seehra,J.S. and Richardson,C.C. (1989) Human immunodeficiency virus 1 reverse transcriptase. Template binding, processivity, strand displacement synthesis and template switching. *J. Biol. Chem.*, **264**, 4669–4678.
46. Yu,H. and Goodman,M.F. (1992) Comparison of HIV-1 and avian myeloblastosis reverse transcriptase fidelity on RNA and DNA templates. *J. Biol. Chem.*, **267**, 10888–10896.
47. Thrall,S.H., Krebs,R., Wohrl,B.M., Cellai,L., Goody,R.S. and Restle,T. (1998) Pre-steady-state kinetic characterization of RNA-primed initiation of transcription by HIV-1 reverse transcriptase and analysis of the transition to a processive DNA-primed polymerization mode. *Biochemistry*, **37**, 13349–13358.
48. Kati,W.M., Johnson,K.A., Jerva,L.F. and Anderson,K.S. (1992) Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.*, **267**, 25988–25997.
49. DeStefano,J.J., Bambara,R.A. and Fay,P.J. (1993) Parameters that influence the binding of human immunodeficiency virus reverse transcriptase to nucleic acid structures. *Biochemistry*, **32**, 6908–6915.
50. Kim,B., Ayran,J.C., Sagar,S.G., Adman,E.T., Fuller,S.M., Tran,N.H. and Horrigan,J. (1999) New human immunodeficiency virus, type 1 reverse transcriptase (HIV-1 RT) mutants with increased fidelity of DNA synthesis. Accuracy, template binding and processivity. *J. Biol. Chem.*, **274**, 27666–27673.
51. Schultz,S.J., Zhang,M., Kelleher,C.D. and Champoux,J.J. (2000) Analysis of plus-strand primer selection, removal and reutilization by retroviral reverse transcriptases. *J. Biol. Chem.*, **275**, 32299–32309.
52. Wisniewski,M., Balakrishnan,M., Palaniappan,C., Fay,P.J. and Bambara,R.A. (2000) The sequential mechanism of HIV reverse transcriptase RNase H. *J. Biol. Chem.*, **275**, 37664–37671.
53. DeStefano,J.J., Critofaro,J.V., Derebail,S., Bohlayer,W.P. and Fritzgerald-Heath,M.J. (2001) Physical mapping of HIV reverse transcriptase to the 5′ end of RNA primers. *J. Biol. Chem.*, **276**, 32515–32521.
54. Pop,M.P. and Biebricher,C.K. (1996) Kinetic analysis of pausing and fidelity of human immunodeficiency virus type 1 reverse transcriptase. *Biochemistry*, **35**, 5054–5062.
55. Pace,C.N. and McGrath,T. (1980) Substrate stabilization of lysozyme to thermal and guanidine hydrochloride denaturation. *J. Biol. Chem.*, **255**, 3862–3865.
56. Brewer,J.M. and Wampler,J.E. (2001) A differential scanning calorimetric study of the effects of metal ions, substrate/product, substrate analogues and chaotropic anions on the thermal denaturation of yeast enolase 1. *Int. J. Biol. Macromol.*, **28**, 213–218.
57. Davies,J.F., Hostomska,Z., Hostomsky,Z., Jordan,S.R. and Matthews,D.A. (1991) Crystal structure of ribonuclease H domain of HIV-1 reverse transcriptase. *Science*, **252**, 88–95.
58. Katayangi,K., Miyagawa,M., Matsushima,M., Ishikawa,M., Kanaya,S., Ikehara,M., Matsuzaki,T. and Morikawa,K. (1990) Three-dimensional structure of ribonuclease H from *E.coli*. *Nature*, **347**, 306–309.
59. Katayangi,K., Okumura,M. and Morikawa,K. (1993) Crystal structure of *Escherichia coli* RNase HI in complex with Mg^{2+} at 2.8 Å resolution: proof for a single $Mg(^{2+})$-binding site. *Proteins*, **17**, 337–346.

60. DeStefano,J.J., Mallaber,L.M., Fay,P.J. and Bambara,R.A. (1994) Quantitative analysis of RNA cleavage during RNA-directed DNA synthesis by human immunodeficiency and avian myeloblastosis virus reverse transcriptase. *Nucleic Acids Res.*, **22**, 3793–3800.
61. Furfine,E.S. and Reardon,J.E. (1991) Reverse transcriptase. RNase H from the human immunodeficiency virus. Relationship of the DNA polymerase and RNA hydrolysis activities. *J. Biol. Chem.*, **266**, 406–412.
62. DeStefano,J.J., Buiser,R.G., Mallaber,L.M., Bambara,R.A. and Fay,P.J. (1991) Human immunodeficiency virus reverse transcriptase displays a partially processive 3′ to 5′ endonuclease activity. *J. Biol. Chem.*, **266**, 24295–24301.
63. Ji,X., Klarmann,G.J. and Preston,B.D. (1996) Effect of human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein on HIV-1 reverse transcriptase activity *in vitro*. *Biochemistry*, **35**, 132–143.
64. Hizi,A., Leis,J.P. and Joklik,W.K. (1977) The RNA-dependent DNA polymerase of avian sarcoma virus B77. Binding of viral and nonviral ribonucleic acids to alpha, beta 2 and alphabeta forms of the enzyme. *J. Biol. Chem.*, **252**, 6878–6884.
65. Modak,M.J. and Marcus,S.L. (1977) Purification and properties of Rauscher leukemia virus DNA polymerase and selective inhibition of mammalian viral reverse transcriptase by inorganic phosphate. *J. Biol. Chem.*, **252**, 11–19.
66. Kacian,D.L. and Myers,J.C. (1976) Synthesis of extensive, possibly complete, DNA copies of poliovirus RNA in high yield and at high specific activities. *Proc. Natl Acad. Sci. USA*, **73**, 2191–2195.
67. VanBeveren,C. and Goulian,M. (1979) Optimal conditions for synthesis of long complementary DNA product with Moloney murine leukemia virus. *J. Virol.*, **30**, 951–954.
68. Skalka,A.M. (1993) Endonuclease activity associated with reverse transcriptase of avian sarcoma-leukosis viruses. In Skalka,A.M. and Goff,S.P. (eds), *Reverse Transcriptases*. CSHL Press, Plainview, NY, pp. 193–204.
69. Nissen-Meyer,J., Raae,A.J. and Nes,I.F. (1981) Effect of ATP on the Friend murine leukemia virus-associated endonuclease activity and the endonuclease activity of the avian myeloblastosis virus RNA-directed DNA polymerase. *J. Biol. Chem.*, **256**, 7985–7989.
70. Fuchs,B., Zhang,K., Rock,M.G., Bolander,M.E. and Sarkar,G. (1999) High temperature cDNA synthesis by AMV reverse transcriptase improves the specificity of PCR. *Mol. Biotechnol.*, **12**, 237–240.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Gerard *et al.*

Appl. No. 09/245,026

Filed: February 5, 1999

For: **Compositions and Methods for Reverse Transcription of Nucleic Acid Molecules**

Art Unit: 1637

Examiner: Strzelecka, T.

Atty. Docket: 0942.4330004/RWE/MTT

Declaration Concerning the Deposited Biological Material

Assistant Commissioner for Patents
Washington, DC 20231

Sir:

I, Alan W. Hammond, on behalf of Invitrogen Corporation, having business concerns at 1600 Faraday Avenue, Carlsbad, California 92008, declare and state as follows:

Escherichia coli DH10B(pDABH-His) was deposited under the terms of the Budapest Treaty on April 15, 1997. This deposit was made at the Agricultural Research Culture Collection (NRRL) International Depositary Authority, 1815 N. University Street, Peoria, Illinois 61604, and given accession number NRRL B-21679. *Escherichia coli* (pDAMVABH-) was deposited under the terms of the Budapest Treaty on June 17, 1997. This deposit was made at the Agricultural Research Culture Collection (NRRL) International Depositary Authority, 1815 N. University Street, Peoria, Illinois 61604, and given accession number NRRL B-21790.

Assurance is hereby given that: (1) all restrictions on the availability to the public of the deposited materials will be irrevocably removed upon the granting of a patent, subject to 37 C.F.R § 1.808(b); (2) the materials have been deposited under conditions such that access to the material will be available during the pendency of the patent application to one determined by the Commissioner to be entitled thereto; and (3) the deposits will be maintained with all of the care

necessary to keep them viable and uncontaminated for a period of at least five years after the most recent request for the furnishing of a sample of the deposit, and in any case, for a period of at least thirty years after the date of deposit or for the enforceable life of the patent, whichever period is longer.

I have read and understood 37 C.F.R. § 10.18(b) and (c).

Respectfully submitted,

Date: October 2, 2002

By: _____

Alan W. Hammond, Esquire
Chief Intellectual Property Counsel
Invitrogen Corporation

::ODMA\MHODMA\SKGF_DC1;56588;1

177

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORGANISMS
FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

TO
Dr. Alan Hammond
Life Technologies, Inc.
9800 Medical Center Drive
Rockville, MD 20850

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT
issued pursuant to Rule 7.1 by the
INTERNATIONAL DEPOSITARY AUTHORITY
identified at the bottom of this page

NAME AND ADDRESS
OF DEPOSITOR

I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR: *Escherichia coli* DH10B(pDABH-His)	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: NRRL B-21679

II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION

Th microorganism identified under I. above was accompanied by:

☐ a scientific description

☒ a proposed taxonomic designation

(Mark with a cross where applicable)

III. RECEIPT AND ACCEPTANCE

This International Depositary Authority accepts the microorganism identified under I. above, which was received by it on April 15, 1997(date of the original deposit)[1]

IV. RECEIPT OF REQUEST FOR CONVERSION

The microorganism identified under I. above was received by this International Depositary Authority on (date of the original deposit) and a request to convert the original deposit to a deposit under the Budapest Treaty was received by it on (date of receipt of request for conversion).

V. INTERNATIONAL DEPOSITARY AUTHORITY	
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority Address: 1815 N. University Street Peoria, Illinois 61604 U.S.A.	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s): Date: 11-14-97

[1] Where Rule 6.4(d) applies, such date is the date on which the status of international depositary authority was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORGANISMS
FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

TO
Life Technologies, Inc.
Attn: Dr. Alan Hammond
P. O. Box 6482
9800 Medical Center Drive
Rockville, MD 20850

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT
issued pursuant to Rule 7.1 by the
INTERNATIONAL DEPOSITARY AUTHORITY
identified at the bottom of this page

NAME AND ADDRESS
OF DEPOSITOR

I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR: *Esch richia coli* (pDAMVABH$^-$)	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: NRRL B-21790

II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION

The microorganism identified under I. above was accompanied by:

☐ a scientific description

☒ a proposed taxonomic designation

(Mark with a cross where applicable)

III. RECEIPT AND ACCEPTANCE

This International Depositary Authority accepts the microorganism identified under I. above, which was received by it on June 17, 1997(date of the original deposit)[1]

IV. RECEIPT OF REQUEST FOR CONVERSION

The microorganism identified under I. above was received by this International Depositary Authority on (date of the original deposit) and a request to convert the original deposit to a deposit under the Budapest Treaty was received by it on (date of receipt of request for conversion).

V. INTERNATIONAL DEPOSITARY AUTHORITY	
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority Address: 1815 N. University Street Peoria, Illinois 61604 U.S.A.	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s): J. L. Snyder Date: 7.5.97

[1] Where Rule 6.4(d) applies, such date is the date on which the status of international depositary authority was acquired.

Certificate Under 37 C.F.R. § 3.73(b)

Applicant/Patent Owner: Gerard *et al.*

Application No./Patent No.: 09/064,057 Filed/Issue Date: April 22, 1998

Entitled: Compositions and Methods for Reverse Transcription of Nucleic Acid Molecules

Invitrogen Corporation, a corporation
(Name of Assignee) (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that it is:

1. [X] the assignee of the entire right, title, and interest, or

2. [ ] an assignee of an undivided part interest

in the patent application/patent identified above by virtue of either:

A. [ ] An Assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the Patent and Trademark Office at Reel _____, Frame ____, or for which a copy thereof is attached.

OR

B. [X] A chain of title from the inventor(s) of the patent application/patent identified above to the current assignee as shown below:

1. From: Gary F. Gerard, Michael D. Smith and Deb K. Chatterjee
To: Life Technologies, Inc.
The document was recorded in the Patent and Trademark Office at Reel 9353, Frame 0181, or for which a copy thereof is attached.

2. From: Life Technologies, Inc. To: Invitrogen Corporation
The document was recorded in the Patent and Trademark Office at Reel _____, Frame _____, or for which a copy thereof is attached.

3. From: _____ To: _____
The document was recorded in the Patent and Trademark Office at Reel _____, Frame _____, or for which a copy thereof is attached.

[ ] Additional documents in the chain of title are listed on a supplemental sheet.

[X] Copies of assignments or other documents in the chain of title are attached.
[NOTE: A separate copy (*i.e.*, the original assignment document or a true copy of the original document) must be submitted to Assignment Division in accordance with 37 CFR Part 3, if the assignment is to be recorded in the records of the PTO. See MPEP 302-302.8]

The undersigned (whose title is supplied below) is empowered to act on behalf of the assignee.

Signature: _____

Name: Alan W. Hammond

Title: Chief Intellectual Property Counsel

Date: July 21, 2003

SRGP_DC1.125726.1